

```

AC Q9R017;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3.
GN ZAP3 OR ZAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RA Misawa K., Nosaka T., Kitamura T.;
RT "A huge nuclear protein rich in proline similar to human hypothetical
RT protein zap3 and zap13."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: Nuclear (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB03168; BAA85182.1;
DR MGD; MGI:1926195; Zap3.
KW Nuclear protein.
FT DOMAIN 15 204 PRO-RICH.
FT DOMAIN 355 473 GLN-RICH.
FT DOMAIN 925 1012 ARG-RICH.
SQ SEQUENCE 1386 AA; 155130 MW; D862F918ED221DF CRC64;

Query Match 5.68; Score 181.5; DB 1; Length 1386;
Best Local Similarity 20.98; Pred. No. 0.3; 266; Indels 237; Gaps 30;
Matches 149; Conservative 61; Mismatches

QY 34 PEDEPAAEAHFLRRHRLKEPPGLASSGSGSPAPSPACGGKGLLPAGAPGQOE 93
DB 148 PESPPVPGSYMPSPSYMPPOPPPSYPPSSAQP-----YLPAPPSPSKPO 196
QY 94 ESMGGSVLPICPP--PATKQAGIGGEPAAGA-----GCSPPR----- 129
DB 197 -----LPPPSISGNKTAIQEPLSGAKNKTAEGKQAPEDPPTMTPEQOQYW 248
QY 130 -----KYAVLPIDTGSIVAAKERTPMAGDKGGAAPATAASDAGPPPLP- 178
DB 249 YRQHLLSLQORTKVLHPGHRKGLVTAADVPEPIKEAPGPAASQVAEPLAAEKPPPLPPN 308
QY 179 --GPPPLAP-----TATAGTLAASGRMKSMKSPSLGGGGS 213
DB 309 EEAAPPPLSPREPSSESESESESEDEARFKQLAKIAAHQAAAHQOQOQOQYV-GEYOY 367
QY 214 GASSQAACLKQILLDLLEQOQOOLQAKKEIEELKSERDTLLARIEMERMOLVKK 273
DB 368 GIMQRTHTLOQIL-----QQYQOVIQHSPIQTP--SLDVQLRHYEQMOOQOFLFQ 417
QY 274 DNEKE-----RHK-LFGYETE-----ERETELSEKI-----K 301
DB 418 DMEREQLMEEOQLHSTPHKQOLOEYEQWKTMOGHMKATDTYQAEKVNSTQYVKSQYLG 477
QY 302 LECQPELSETSQTLPRKPFSCGRSGKHKRKSFPGETERTKTPVKLAPEFSKVKTKPKH 361
DB 478 MAMPPEFVPSQMPPLP-----TMPPVPLP--SLPPVVPRA 514
QY 362 SPIKEPCGSLSTVCKRELRSQETPEKPRS-----SVDPRLSTPQKGPSTHREKA 415
DB 515 LPTSIPTPPGMPVPVMPPLPTSVPPGMPPLSLSSGPPVLPPLPALSAGSPVLPALP 574
QY 416 FSSEIEDLPYLTSTTEMYLCRWHPPP-----SPLPLRESPPKE--ETV 457

```

```

DB 575 GCPPIILPLPLSSA-----TPPGIIPPGAPQGMPPQATLAPPPASSQNSQIPEK 626
QY 458 ARCLMPSVA-GFTSVLAVPSMRDHSVEPLRDNPDLLENLD-DSVEFSRHAHKLDEK 515
DB 627 RQALLPLPVSEGSTP-----PS-----PYHPPOSEOVNSKPLNPFSSQG--LGES 672
QY 516 RRRKWDIQRIREORIL-----ORQLRMVKKKGIOSEPEVTS 553
DB 673 SALSQSTIAKQDPVAVSGGLLADPPKGSFLEGRRGREKQEQQLKDG---SEFQMD 729
QY 554 FPEPDVESLMTTP-----FLPVAFGRP-----LPKLPQNFELP 590
DB 730 HLPDPD-----SLQNPSPRGMYPPPGSYRPPPMGPKPGSIVRSAPPARSSIP 779

RESULT 11
SHK1_RAT
ID SHK1_RAT STANDARD; PRT; 2167 AA.
AC Q9WV48; Q9WU13; Q9WUE8; Q9QZ48;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S63 and multiple ankyrin repeat domain protein 1 (Shank1) (GKAP/SAPAP
DE interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
DE interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RP DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Tao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synmon, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein."
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRATIN-Sprague-Dawley;
RX MEDLINE=99360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin."
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor CL1 interacts directly with proteins
RT of the Shank family."
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RP DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
RT genes, alternative splicing, and differential expression in brain and
RT development."
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoernck H.-H., Baechner D., Richter D., Kreinbamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
RT multidomain proteins present in human and rodent brain."

```

GN SHANK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.
RC TISSUE=Fetal brain, Hippocampus, and Thalamus;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoehn H.-H., Baecher D., Richter D., Krielenkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [2]
RP REVIEW.
RX PubMed=10806096;
RA Sheng M., Kim E.;
RT "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
density (PSD) of excitatory synapses that interconnects receptors
of the postsynaptic membrane including NMDA-type and metabotropic
glutamate receptors via complexes with GKAP/PSD-95 and Homer,
respectively, and the actin-based cytoskeleton. May play a role in
the structural and functional organization of the dendritic spine
and synaptic junction.
CC -1- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with
SPAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2. (By similarity).
Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
neuronal cells (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1/a (shown here), 2/b and 3; are
produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in brain particularly in the
amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
seems to be expressed ubiquitously.
CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DR EMBL: AF163302; AADA5121.1; -
DR EMBL: AF226728; AAF35887.1; -
DR HSSP: P06241; 1SHF.
DR MIM: 604999; -
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 7.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00536; SAM; 1.
DR Pfam: PF00018; SH3; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00248; ANK; 3.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00454; SAM; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50088; ANK_REPEAT; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50106; PDZ; 1.
DR PROSITE: PS50105; SAM_DOMAIN; 1.

DR PROSITE: PS50002; SH3; 1.
KW ANK Repeat; SH3 domain; Repeat; Alternative splicing.
FT REPEAT 212 245 ANK 1.
FT REPEAT 246 278 ANK 2.
FT REPEAT 279 312 ANK 3.
FT REPEAT 313 345 ANK 4.
FT REPEAT 346 378 ANK 5.
FT REPEAT 379 395 ANK 6.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2098 2161 SAM.
FT DOMAIN 1002 1007 POLY-HIS.
FT DOMAIN 1014 1019 POLY-HIS.
FT DOMAIN 1189 1195 POLY-GLY.
FT DOMAIN 1709 1717 POLY-GLY.
FT DOMAIN 1844 1854 POLY-PRO.
FT DOMAIN 1896 1902 POLY-GLY.
FT DOMAIN 1970 1979 POLY-SER.
FT VARSPIC 1 613 MISSING (IN ISOFORM 2).
FT VARSPIC 614 654 MISSING (IN ISOFORM 3).
FT VARSPIC 646 654 MISSING (IN ISOFORM 3).
FT SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;
SQ
Query Match Best Local Similarity 22.5%; Score 182.5; DB 1; Length 2161;
Matches 128; Conservative 44; Mismatches 191; Indels 207; Gaps 28;
OY 32 GCPDEPCGAENHPLPRKRLKEPPPLASQSGSPASP-----ACGGCKRG 80
DB 1030 GSPDDP-----PRLAGPQPSLRGRRGGSPPTPAPSPSHSGAGGGGSSQG 1079
OY 81 -----LLLPAGAA-----PCQEESSMGVPLPCPPATQAGIGEPAAAGACSPRP 129
DB 1080 PALKRIQLPPRAASAMVVPARGRGK-----PLVQTKVEGP-QKGGGLPPAP 1130
OY 130 KYQAVLPLOTGLVAAKE-----PTPWAG 154
DB 1131 SPSSPASPQPPPAVAAPSEKNSIPITIIKAPSTSSGSSGSGSTAEPPTPQDEPTGG 1190
OY 155 DKGGAAPA-ATASDPAGPPLPLP-----GPPPIATYATAGT-----LAASEGRW-KS 201
DB 1191 GGGGSSPSPAPAMSPVPPSPSPVPTPASPSGPATIDTSGFGAALVGAAAREGQWNEA 1250
OY 202 MKRSPL-----GGGSGGASQAACLKQILLDLDLEQQOQQLQAKEREIEELK 251
DB 1251 RRRSTLPLSTAGDEGDDGLGTGA-----APGRRLRSK 1286
OY 252 SERDTLLARIEMERRMQLVKKDEKERHKLFGYTEERETELSEKIKLECOPELSET 311
DB 1287 SIDEGMFS-----AEYRLLESAGSGAG-----YCGGAGSRAVYGGGGS-----SAF 1329
OY 312 SQTLPKPKPSCGSGSKGKRRKSPFG-----STER-----KTVYKKLAPFESVYK 355
DB 1330 TSFLPRLPLVPLLTGKALDPASPGLALAAERALKESSEGGAQPPRPSPPEYEAAP 1389
OY 356 TKTPKHSP-IKEEPC-----GSLSETVCKRE-----LRSOE-----TPEKRRSV--DTP 397
DB 1390 PTPHHSHAHHEPVLRLMGASPPDPARRELGYRAGLSQSEKSLPASPAAARRSLHMLP 1449
OY 398 PR-----LSTPQKGPSTHPK-----EKAFFSEIEDLPYLSTTEMYLCRMHQ----- 438
DB 1450 PTAQGVGPLLQLTGTEPPAPHPGVSKPKPMSAARPEPRLP-LHVRFLNCGPRAPVYISGR 1508
OY 439 -----PPSP--LPLRESSPKKEE 455
DB 1509 GPPSEDPGVPSPSPRSRVSPPSPSPRASE 1538
RESULT 10
ZAP3_MOUSE STANDARD: PRT; 1386 AA.
ID ZAP3_MOUSE


```

Db 146 GTAESGEAARAIPRE-----LPVKSSSLPEIKPEHKGFLPNHFNAGRGSRSELG 200
OY 159 --ASPATASD-----PAGPPL-----PL 177
Db 201 RAAGRGASDADGKLRNHFVGRSTVTKTLPAKPKHVELNLTGKPNLDSLGEMHNF 260
OY 178 PGPPPLATPATAGTLAASEGRWKSRLGGGGSGASQAACLKQIILLDLLEQO 237
Db 261 SOPVHKGMTATKISLFENKRTKNSPRHNDIRGPRNTPASSKI-----FVGRAK 308
OY 238 QOLAKKEKEIEELK-----SERDTLLARIEEMERMOVLKDD-----274
Db 309 LNLAKKAKEMQPEKVKMPSNGVLKETAIEKVVSEELIPATRGMDGSENQA 368
OY 275 -----NEKER-----HKLFGQYETEEREETELSKILECOP 306
Db 369 LGPQONODKADVQDAGCLSEPVASALIPVKDHL-----EKEDSPAADSKSLVE- 421
OY 307 ELSETSQTLP-----PKPFCGSRGKGHKRSPPGSTERKTPYKTLAPFSK 353
Db 422 NVTDAQDIPPTVDKDLPTAMPKQHTFSDSQSPASSPGSLSLAP---AP--GD 475
OY 354 VKTKPKHSPTKEECGSL-----SETVCKRELRSGETEKPSSV--DTPPLST---PQ 404
Db 476 VPKDTCVOSPSSFPCTDLKVENHKGCVLPVSRQNNKEMPLELGGTTPLSTERGPE 535
OY 405 KGPSTHPKE-----KAFSEIEDLPYLST-----TEWLCRMHQPSPPLPLRESS 450
Db 536 AVGSECFRVLVQVNSFLPVSTODVSSQVTPESSEVREVLPCHSNEPVSVASCA 595
OY 451 PKKEETVA-----RCLMPS-----VAGETSVLAVPSW 478
Db 596 PPOEEVLNENSHCTAELAAKSGPOVIPPASEKTLPIQAQSGSRTPLMAESSPTNPS 655
OY 479 RDHSVEPLR-----DPNPSDL-----NDDSVFSRHKLELDEKRRKMDQRIRE 527
Db 656 GNLHATPQRPDQVTYNGDSPASLTINISAGSDSVFSSS-----DKKETE 702
OY 528 ORILQRLQ-----LRMYKKKGIOESEPEVTSPFPEDVESLMTFLP-LVAFG-----R 577
Db 703 --IIQMDSAYCMPPKRRKAKMPNSP--APHFAMPPIHEDHLEKVFDFKVFGLGKKKE 758
OY 578 PLPKLTPQNFELPWIDERSRCH 599
Db 759 SQEWSPALHLMQNDTKSLR 780

RESULT 7
WASP_MOUSE STANDARD: PRT; 520 AA.
ID 70315;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein homolog (WASP).
GN WAS OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c;
RX MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.:
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sfi) mutation on
RT the X chromosome."
RL Genomics 29:471-477(1995).
CC -I- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION
CC (BY SIMILARITY).

```

```

CC -I- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -I- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC EMBL: U54788; AAC52556.1; -.
CC MGD: MGI:105059; Was.
CC InterPro: IPR000095; Pakbox/Rhodning.
CC InterPro: IPR000697; Ranbpl_WASP.
CC InterPro: IPR001960; WH1.
CC InterPro: IPR003124; WH2.
CC Pfam: PF00568; WH1; 1.
CC Pfam: PF00786; PBD; 1.
CC Pfam: PF02205; WH2; 1.
CC SMART: SM00285; PBD; 1.
CC SMART: SM00461; WH1; 1.
CC SMART: SM00246; WH2; 1.
CC PROSITE: PS50108; CRIB; 1.
CC Repeat.
CC FT DOMAIN 41 147 WH1.
CC FT 240 253 CRIB.
CC REPEAT 354 363 GRSGLPPXP MOTIF 1.
CC FT 393 402 GRSGLPPXP MOTIF 2.
CC FT DOMAIN 162 167 POLY-PRO.
CC FT DOMAIN 314 321 POLY-PRO.
CC FT DOMAIN 324 341 POLY-GLY.
CC FT DOMAIN 368 373 POLY-PRO.
CC FT 376 379 POLY-PRO.
CC FT DOMAIN 384 390 POLY-PRO.
CC FT 397 403 POLY-PRO.
CC FT DOMAIN 408 424 POLY-PRO.
CC FT 503 520 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 520 520 ASP/GLU-RICH (ACIDIC).
CC SQ SEQUENCE 520 AA; 54191 MW; 9C223733C9FC68 CRC64;

Query Match
Best Local Similarity 26.1%; Pred. NO. 0.04;
Matches 80; Conservative 29; Mismatches 106; Indels 91; Gaps 12;

OY 15 AGNPEQRLDYERAAALGPEDEPGAAR--HFLPRHKIKERPGPLASSOGCGSPAPSPA 72
Db 276 AGISEAQLTDAETSKLITDFLEDQGLEAVRQEMRQEBPLPPPPCRGGGGGGGGGG 335
OY 73 GCGGGRGLLPAGAAPQOEESMGSVLP-----CPPATKQAGIGEPAAAGAGCS 126
Db 336 GGGGGGQPLRPVVS-----NKRSGPLPPVPGAGAPPTTPG-----PPPRGRGP 384
OY 127 PRPKYQAVLPQTGSLVAANEKPTPMADKG-----GAASPAATASDPAGPPLPLPG 179
Db 385 PPP-----PPPATGRSGPPPLPGAGGPPAPPPPPPPPPPPCG 425
OY 160 -----PPPLPATATAGTLAASEGRWKSRLGGGGSGASQAACLKQIILLDLLEQ 235
Db 426 SGPAPPPLPPPPVSGSPAP-----GGGRGA-----LLDQ 455
OY 236 QOQOQO-----AKKEIEELKSERDTLL--ARIPMERMOVLKONEKERHKLFGYE 287
Db 456 IROGIQLKKTALENSVOQPPAQOSEGLVGLAHVYMKRSRVTHSSDEGD-----QICE 511
OY 288 TEEREE 293
Db 512 DEEDE 517

RESULT 8
WASP_MOUSE

```



```

FT VARSPLIC 532 582 VSANSSVITKTRFRMEDPPKPKVEVSPVLRPPITTSPP
FT VPOEPAPVPS -> PLSOSLIPMTLDSLSIGOMAPPTIS
FT ACIDSPPLMSPLLRPCPLTGLQ (IN ISOFORM
FT VARSPLIC 583 2715 MISSING (IN ISOFORM TRUNCATED).
FT CONFLICT 834 834 K -> E (IN REF. 6).
FT CONFLICT 941 941 S -> Y (IN REF. 6).
FT CONFLICT 1317 1317 E -> Q (IN REF. 6).
FT CONFLICT 1362 1362 H -> Y (IN REF. 6).
FT CONFLICT 1438 1438 D -> N (IN REF. 6).
FT CONFLICT 1918 1920 PLA -> GTR (IN REF. 5; AAH09337).
FT CONFLICT 2541 2543 DEE -> ARG (IN REF. 5; AAH07353).
FT CONFLICT 2622 2622 D -> H (IN REF. 6).
SQ SEQUENCE 2715 AA: 293511 MW: C0615B981BBEB7BF CRC64:

Query Match 6.1% Score 196.5; DB 1; Length 2715;
Best Local Similarity 21.0%; Pred. No. 0.13; Indels 285; Gaps 36;
Matches 158; Conservative 61; Mismatches 248;

28 AAALGPEDEPGAHAHFLPRHRLKEPGLPLASSOGSPAPSPAGCGKGR-----79
3 AAAGGCGPGSGSARGRF-----PGRPRGAGGG-----GGGGRNGAERYVA 46
80 -GLLPAGAPGQ-----OESWGGSVP 101
47 LRRCGATGPGGAPGEDTALLRLRLRLRLRLMAGPRVGRGRGRGNGWGPS--104
102 LPCPPATKQAGIGGEPAAAG-----AGCSPR-----P 129
105 RGVPEEESDGEDEEFOGFHSDVPASSLRSALRSORGNAPRGKHKHTPLPP 164
130 KYQAVLDIQGSLVAAAKEPTPMAGDK-----GAASPAATASDPAGP- 172
165 RLADVAP-----TPKTPARKRGEGETERMVQATLRLRAQAQAPRSRACEPTPR 217
173 -PP-----LPPL-GPPPLAFTATAGTLASEGRKMSMR- 203
218 RSRGRPPGRAPGPCRRKQAAVVAEAAVTTPKPPPPPPV-----VKHOTGSMCKEG 271
204 -----KSPLGCGSGSGAS-----SOAACTKQIILLLODL-IEQQQ 238
272 PGCPGTPRRGGGSSRGGRGRGCGGLPFTVKEVSA-----KVKMGSLSLGSGG 328
239 QLAKEKEIEBELKSERDTLARI-----ERMERRMQLVKKDKNEKHKLFQGYET 288
329 QGQHEE-----SMQVPPORVSGSGGSCWKQKOKLDEDEEKKKEEKKKEGEEK 381
289 EEBETELSEKIKLECPPELSETQTLPPKPFSCGSGKSGHKRSKSGSTERTKTPV- KKL 347
382 EERAVAE-----EMMPAAKEKEAKLPPLTPPAPSPPLPP--STSPPLPLCPP 432
348 APESKVKTKPKPKSPIKEE-----PCGSLSETVCKRELRSQETPEK-----PRSS 393
433 PPRVSPPLPSPPPPPOEEOEESPVPVATCSKRGKRRPLPLTSORAEARAAGPBG 492
394 VDPPEPLST-----POKGPSTHREKAKFSEIED--LPYLSTTEMYLCR-----WHOPPP 441
493 SPPTPTSTATGCPEDSPVAPAKSTFELKNIRQFIVPVASBSRVIKTPRRMDPDP 552
442 SPLPLRESSPKKETVARCLMPSSVAGETSVLAVPSMRDHSVEPLRDPNSDLENDOS 501
553 KP-PKEVESP-----VLRRPITTSPPVPOEPAPVPS-----PPRAPTPPS-----591
502 VFSKRAHLELDEKRRKRWIDIORIREORILORLQRLMYKKKKIQEESPEVTSFPEPDDV 561
592 -TPVPLPERRS-----ILREPTFKMTSL--TRELPPLPPPA--PPPPPA 630
562 ESLMTFFLPVAVAGRPPL---PKLTPQNEEL 589
631 PS---PPAPATSSRRPLLLRAPOFTPSEHL 659

```

```

AIM1_HUMAN
ID AIM1_HUMAN STANDARD; PRT; 1723 AA.
AC 09Y4K1: 000296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Absent in melanoma 1 protein.
GN AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Liver.
RX MEDLINE-97250519; PubMed-9096375;
RA Ray M.E., Wistow G., Su Y.A., Meltzer P.S., Trent J.M.;
RT "AIM1, a novel non-lens member of the beta-gamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -1- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U83116; AAB53792.1; -.
DR EMBL: U83115; AAB53791.1; -.
DR HSSP: P02526; IGAM.
DR Genew: HGNC:356; AIM1.
DR MIM: 601797; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00030; Crystallin; 6.
DR Pfam: PF00652; Ricin_B_lectin; 3.
DR PRINTS: PR01367; BGCRYSTALLIN.
DR SMART: SM00458; RICIN; 1.
DR SMART: SM00247; XTALDg; 6.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 7.
DR PROSITE: PS0231; RICIN_B_LECTIN; 1.
KW Repeat; Lectin.
FT DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.
FT DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.
FT DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.
FT DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.
FT DOMAIN 1405 1497 CRYSTALLIN BETA/GAMMA-LIKE 5.
FT DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.
FT DOMAIN 1586 1719 RICIN B-TYPE LECTIN.
SQ SEQUENCE 1723 AA: 188646 MW: 7550F681A627B09 CRC64;

Query Match 6.0% Score 193.5; DB 1; Length 1723;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 164; Conservative 102; Mismatches 269; Indels 267; Gaps 38;

6 AVEFAAAP-----AGNPEQRLDVERAAALG-----PPDEPGAHAHFLPRHRLKEP 55
38 AVFDEVAAPNADSNASAEKKVKSPPRALDGVASASAPSPKSPGKGL- RGSERDK 96
56 GPPLASSOGSPAPSPAGCGGKGRGL-LTPAGAAPGQOEESWGGSVP--LPCPPATKQA 112
97 QP-----PASSPTKRGRSRALAEVAVAPASGRARA-KESPPKRRVPDSSPYTK-- 145
113 GIGGEPAAGAGCGSPRKYAVLPIQTGSLVAAAKE-----PTPWAG-DKGG----- 158

```

OY 436 WHOP-----PPSPLEPRESSPKKEETVARCLMPSSVA-GETSVLAVPSMR 479
 DB 766 LROPHNRIMLTPTPOIQINLPQPVV-----VAPVLPCKALKSAVSAQAAAOK 815
 OY 480 DHSVEP-----LRDPPSLLENLDDSVFSKRHAKI-----ELDEKRRKRWIDQIRREORI 530
 DB 816 NKLKEPGGSGFRDDDDINDVASMAGVNLSEESARILATNSELVGLTRSCDEFFLLQAP 875
 OY 531 LQRLQLRMVKKKGIOSEPEVTSF 554
 DB 876 LQRLLEIGKKHGIELHPDVVSY 899

RESULT 5
 TRX2_HUMAN STANDARD. PRT: 2715 AA
 ID TRX2_HUMAN Q9JUN6; Q9JUN25; Q95836; Q91669; Q91668; O15022; Q96GP2; Q96IP3;
 AC Q9JUN6; Q9JUN25; Q95836; Q91669; Q91668; O15022; Q96GP2; Q96IP3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).
 GN TRX2 OR HX2 OR ML2 OR ML4 OR KIAA0304.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RE SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Angard P.O., Valvatne H., Jeanmougin F., Adamson A.,
 RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
 RA Lamerdin J., Chambon P., Lossen R., Stewart A., Aasland R.;
 RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin
 RT regulators which contain PHD fingers and SET domains";
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RE SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhart-Schultz K.,
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Ganes J., Dangnan L.,
 RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
 RA Carraro A.V.;
 RT "Sequence analysis of a 1 Mb region in human 19q13.1";
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RE SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).
 RP TISSUE-Testis, and Leukocyte;
 RX MEDLINE=20105772; PubMed=10637508;
 RA Hiedeman D.G., Chin S.-F., Mulertis M., Batley S.J., Collins V.P.,
 RA Wiedeman L.M., Aparicio S., Caldas C.;
 RT "ML2, the second human homolog of the Drosophila trithorax gene, maps
 RT to 19q13.1 and is amplified in solid tumor cell lines";
 RL Oncogene 18:7975-7984(1999).
 RN [4]
 RE SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).
 RP TISSUE-Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [5]
 RE SEQUENCE OF 1918-2715 FROM N.A.
 RP TISSUE-Brain, and Skin;
 RA Strausberg R.;
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RE PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).
 RP TISSUE-Placenta, and Bone marrow;
 RX MEDLINE=99339983; PubMed=10409430;
 RA Fitzgerald K.T., Diaz M.O.;
 RT "ML2: A new mammalian member of the trx/MLL family of genes.";

RL Genomics 59:187-192(1999).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
 CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
 CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL
 CC BLOOD LYMPHOCYTES, AND PLACENTA.
 CC -1- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AJ007041; CAB45385.1; -
 DR EMBL; AD000671; -; NOT_ANNOTATED_CDS.
 DR EMBL; AB002302; BAA20763.2; -
 DR EMBL; AF186605; AAD56420.1; -
 DR EMBL; AF104918; AAD17932.1; -
 DR EMBL; AF105279; AAD26113.1; -
 DR EMBL; BC009337; AAH09337.1; -
 DR EMBL; BC007353; AAH07353.1; -
 DR EMBL; AF105280; AAD26112.1; -
 DR MIM; 606834; -
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR003889; Fitch_C.
 DR InterPro; IPR003888; Fitch_N.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR002857; Znf_CXXC.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_Fing.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR Pfam; PF02008; zf-CXXC; 1.
 DR SMART; SM00384; AT_hook; 1.
 DR SMART; SM00542; FYRN; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS00280; SET; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 3.
 DR PROSITE; PS50016; ZF_PHD_2; 3.
 KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
 KW Transcription regulation; Alternative splicing; Repeat.
 FT DNA_BIND 37 44
 FT DNA_BIND 110 117
 FT DNA_BIND 357 365
 FT ZN_FING 959 1005
 FT ZN_FING 1201 1252
 FT ZN_FING 1249 1303
 FT ZN_FING 1335 1396
 FT DOMAIN 1449 1471
 FT DOMAIN 1486 1491
 FT DOMAIN 2586 2715
 FT DOMAIN 2715 2715
 FT DOMAIN 26 37
 FT DOMAIN 248 255
 FT DOMAIN 362 398
 FT DOMAIN 402 771
 FT DOMAIN 808 812
 FT DOMAIN 1963 1970
 FT DOMAIN 2251 2259
 FT DOMAIN

RESULT 4
T2D3_HUMAN STANDARD: PRT; 1083 AA.
AC 000268: 099721: 09BX42; 09BR40;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII-130) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_Taxid=9606;
RX MEDLINE=97336072; PubMed=9192867;
RA Mensus G., May M., Carre L., Chambon P., Davidson I.;
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
of the retinoic acid, vitamin D3, and thyroid hormone receptors in
mammalian cells.";
RL Genes Dev. 11:1381-1395(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagge C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasialho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McMay K., McMurtry A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symcote N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97098442; PubMed=8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIID
complex: hTAFII130 and hTAFII100.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y11354: CAAT2189.1; -
DR EMBL: AL137077: CAC36006.1; -
DR EMBL: AL109911: CAC22312.2; -
DR EMBL: U75308: AAC50901.1; -
DR TRANSFAC: T02328; -
DR Genew: HGNC:11537; TAF4.
DR MIM: 601796; -
DR InterPro: IPR003894; TAF_hom.
DR SMART: SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein.
FT DOMAIN 39
FT DOMAIN 52
FT DOMAIN 98
FT DOMAIN 142
FT DOMAIN 268
FT DOMAIN 331
FT DOMAIN 680
FT DOMAIN 808
FT DOMAIN 828
FT DOMAIN 105
FT CONFLICT 136
FT CONFLICT 185
FT CONFLICT 233
FT CONFLICT 293
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;
Query Match 6.1%; Score 197; DB 1; Length 1083;
Best Local Similarity 20.8%; Pred. No. 0.051;
Matches 155; Conservative 75; Mismatches 262; Indels 252; Gaps 30;
11 AAAPA-----GCNPGQRDYERRAALGCPDEPPCAAEHFLPRHKKLKEPPPLASSGCGS 66
208 AAAPAVSLVNGPAPALLPLPKPAPAGTVIGTPPVGAADPADPADSPAPAPAPAPAAA 267
67 PAPPAGCGGKGRGLLPAG---AAP-----GQGEESMGVSPLCPPTATQAGIG 117
268 PPPPPAPALPAPLPAGPPAPPPAPPAAPVPPAPPAANGSGAAPAPA-PAAAGPACVSG 326
118 PAAAGACSPRPKYQAVLPYIGTGSVLAARKEPTPMAGDKGAASPAATAS----- 167
327 PGCAAAAPAPGKASPKR---VVOAA-----PPAAQTAAAGPASTASAVIGPTMG 379
168 ---DPAPPLPLPAGPPPLAPPTAGTGLASBGRKWSMRKSPGGGSGSASSQAACLK- 223
380 ALDSPAAPVP-PAPGPTGLPKGAAGAVT-----GSLSTPTATTSIGIRATLPVYLAP 432
224 -----OILLQLDLIPQOOQOOLQAKER----- 245
433 RLPPPPNPINIQFPLPGMVLVRSNGGLMIPQALNOMQAQNAHQPTMAPRAPAT 492
246 -----EIEELKSERDTLAP-----TERERRRQVYKKNEKER----- 279
493 PPSAPVQISTVQAPGPIIAROVTPPTTIKQVSAQTQVPSATLRSGVPOVLVIG 552
280 -----HKLPGYETEEHEHELSKILKEOPELSEHSQTLPPKPF 321
553 AAQTASIGLTAVQGTGPQTPVPGATTSSAAETMENVA-KCKNFSL---TL---IK 603
322 CGRSGGKHKRKSPEFGSTERTPKVKKL-----APESKYKTKTPKSPKEPCGS 371
604 LASSGK-----QSTETANVKELVONLDDKIEADPTSLYKELNSP---QP--- 649
372 LSETVCKREL-----RSQETPERKSSVDT----- 396
650 YLVPFLKRLSPALROLTPDSAAFLOQSOQPPPTSQATALTAAVAVLSSSVQRTAGKTA 709
397 -----PPLSTPPO-----KGPST-----HKERAFSEIEDPLSTTEWYLCR 435
710 TVTSAIQPVLSTLQTPGVGVGKOGPTPLVIOQPPKPALIRP---PQVTLIQDPMA 765

Db 917 - KEDAKPEKTEVAK--KEPDADA 938

RESULT 3

NFH_RAT

ID NFH_RAT STANDARD: PRT: 831 AA.

AC P1684; 063368;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Neurofilament triplet H protein (200 kDa neurofilament protein)

DE (NEFH OR NFH).

GN NEFH OR NFH.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OC NCBI_TaxID=10116.

XX [1]

XX TISSUE=Brain;

XX MEDLINE=89065087; PubMed=3143606;

XX Breen K.C., Robinson P.A., Wilson D., Anderson B.H.;

XX "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).

XX Identification of putative phosphorylation sites."

XX FEBS Lett. 241:213-218(1988).

XX [2]

XX SEQUENCE OF 37-831 FROM N.A.

XX MEDLINE=88309090; PubMed=2457365;

XX Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,

XX Jolles P.;

XX "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and

XX in situ detection."

XX Blochem. Biophys. Res. Commun. 154:1099-1106(1988).

XX [3]

XX SEQUENCE OF 1-89 AND 243-313 FROM N.A.

XX MEDLINE=87080760; PubMed=2878828;

XX Robinson P.A., Wilson D., Anderson B.H.;

XX "Isolation of a cDNA for the rat heavy neurofilament polypeptide

XX (NF-H)."

XX FEBS Lett. 209:203-205(1986).

XX [4]

XX SEQUENCE OF 318-831 FROM N.A.

XX MEDLINE=89184647; PubMed=2928342;

XX Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,

XX Smolowitz M., Carroll Z., Emanuel B.S., Brettner J., Rubin L.;

XX "Cloning of a cDNA encoding the rat high molecular weight

XX neurofilament peptide (NF-H): developmental and tissue expression in

XX the rat, and mapping of its human homologue to chromosomes 1 and

XX 22."

XX Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).

XX [5]

XX FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,

XX AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.

XX NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT

XX OBSERVED BY THE TWO SMALLER NF PROTEINS.

XX -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS

XX PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS

XX THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF

XX INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE

XX OF AXONAL CALIBER.

XX -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING

XX OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

XX LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND

XX COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.

XX -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

XX -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783

XX ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

----- THIS SWISS-PROT entry is copyright. It is produced through a collaboration

----- between the Swiss Institute of Bioinformatics and the EMBL outstation -

----- the European Bioinformatics Institute. There are no restrictions on its

----- use by non-profit institutions as long as its content is in no way

----- modified and this statement is not removed. Usage by and for commercial

----- entities requires a license agreement (see <http://www.isb.ch/announce/>)

or send an email to license@isb-sib.ch.

CC

DR EMBL: M37227; AAA41693.1; ALT_FRAME.

DR EMBL: X13804; CAA32038.1; ALT_FRAME.

DR EMBL: M21864; AAA41695.1; -.

DR EMBL: J04517; AAA41692.1; -.

DR PIR: A30796; A30796.

DR PIR: A25649; A25649.

DR PIR: B25649; B25649.

DR PIR: S02003; S02003.

DR InterPro: IPR001664; IF.

DR Pfam: PF00038; I1.

DR ProSite: PS00226; IF. 1.

KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation; Repeat.

FT NON_TER 1

FT DOMAIN 1

FT CONFLICT 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.

FT CONFLICT 164 164 L -> I (IN REF. 2).

FT CONFLICT 185 185 I -> S (IN REF. 2).

FT CONFLICT 193 193 L -> T (IN REF. 2).

FT CONFLICT 199 199 M -> T (IN REF. 2).

FT CONFLICT 346 346 K -> N (IN REF. 1).

FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).

FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).

FT CONFLICT 485 485 P -> S (IN REF. 2).

FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).

FT CONFLICT 591 591 A -> V (IN REF. 2 AND 4).

FT CONFLICT 727 727 A -> T (IN REF. 4).

FT CONFLICT 757 759 AAP -> GST (IN REF. 4).

FT CONFLICT 769 769 T -> L (IN REF. 2).

FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).

SO SEQUENCE 831 AA; 89486 MW; 180973C3F13EF768 CRC64;

Query Match 6.1%; Score 197.5; DB 1; Length 831;

Best Local Similarity 21.6%; Pred. No. 0.038;

Matches 116; Conservative 89; Mismatches 239; Indels 93; Gaps 22;

QY 11 A A A P A G N P E O R L Y E A A L G P E D F G A E A H L P R H R K - - - L K E P G P L A S S O G S 66

DB 315 A K S P A - - - - - E V K S P A E V K S P A E A K S P A E A K - S P A E V K S P A T V K S P E A K S P A E V K S 365

QY 67 P A - - P S P A G C G G K R G L L P P G A A P G Q Q E E S W G S V P L P C P P P T K Q A G I G E - P A A A G A 123

DB 366 P A E V K S P A E A K S P A E A K S P A S V K S P E A K S P A E A K S P A T V K S P E A K S P A E V K S 425

QY 124 G C S P R K Q A V L P I O T G S L V A A K E P P W A G D G G A A S P A T A S D P A G P P L P G P P L 183

DB 426 P V T V K S P A E A K S P V E V K S - P A S V K S P S E - A K S P G A K S P A E A K S - - - - - P V A A S P A 475

QY 184 A P T A G T L A S E G R K M K S M R K S P L G G G G S G A S S O A C L Q I L L Q D L D I B Q Q Q O L O A K 243

DB 476 E A K S P A G A K P P A E A K S P A E A K S P A E A K S P A E A K S P V - - - E V K S P E A K S P V E G 532

QY 244 E K E I E L K S E D T L A R I E R K R Q L Y K N E K E R H L F G Y E T E E E E L S E K I L E 303

DB 533 A K S I A E A K S P E A K - - - - - K S P V E E I K P P A E V S P E A K S P M R E A K - - 574

QY 304 C O P E L S E S Q L P P - - K P F S G R S G K G H R K S P S G S T E R - T P P K L L A P E R S K V K T Y P K 360

DB 575 - S P E K A K L D V K S P E A K - - - - P A K E A K R A D I R S P Q V A S P A K E A K S E K E T R E K 629

QY 409 T H - P R K A F S S E I D L P Y L S T T E Y L C R W H O P P S P L P L R - E S S P K K E E T V A R C L M P S S V 466

DB 690 T E K P R D S G E A K K E - - - - A K E R K A A P E E E T P A K L G V K E A K P K E A D A K A K E S K P 744

QY 467 A - - - - - G E T S V L A P S M R D H S V E P L R D N P S D L E N D S V F S R H A K L E I D E K R R K 518

DB 745 S E K E P K K E E V P A P E K D T E E K T E S - - - - - K R E E P K M E A K A K E 787

FT CONFLICT 492 492 L -> S (IN REF. 1).
 FT COINCIDENT 670 670 I -> M (IN REF. 1).
 SQ SEQUENCE 1039 AA; 117412 MM; 4759E9B5E6E9F14 CRC64;

Query Match 6.58; Score 211; DB 1; Length 1039;
 Best Local Similarity 20.38; Pred. No. 0.012;
 Matches 96; Conservative 91; Mismatches 171; Indels 114; Gaps 18;

181 PPLAPYTAGTGLAASGRKMSKMRKPLGGGSGGASQAAACLOIIL-----LQDLIE 234
 569 PKHLPAVAKVAKVKTSSRESTLEKRA-----MTADKDAPOAKVIANHSTKTQIDPVK 621
 235 QQQOQLOAKKEI-----EELKSRDT-LIARIERMRRLVAKDKERHKL 282
 622 TORLOVKIROYEMHPDMRTGSSAPSDIRKQNVDPSTPTKTSKSLVNDKKTSET 681
 283 FQGVETREERET---ELSEKIKLECOPELSETSTLP-----KPPSCGSGK 327
 682 SOSPDQETIVEYRKRAEHLKKEILLSHSSQVTLKIRERVAATNLITPPP----- 733
 328 GHRKSPFGSTERKTPYKLAPEFSKYKTRPKH---SPIKEPFGSLSETVCKRELRSQ 384
 734 ---SAVST---TTPATPTPTTPPGSTPGAAVATSSMDQELSAKSKAKAEQI--- 783
 385 ETPEKPRSSVDTPPRUSTPOK---GPTHPKEKAFSSIED---LPVISTT--- 429
 784 ATPLTPQSSSVSSTSTIRKTLNCSPHYTSKATARSGLQSFRTATPEYSTRTWEDQ 843
 430 -----EMYLCHRMHPPSPPLPRESKPEETVAROLMSSVAGETSVLAIVPMRHSV 483
 844 EPHCDNEFL-----EEA-----DELLADNPSSLEIPKWRDPV 876
 484 EPLRDPNPSSDLENLDVSFKRAKLELDEKRRMDIORIRQRILOLQRLMYKKG 543
 877 PPSDRKIDTEL---LSQATFERRRQKQVYKDEVDKCKDAYKMEQILDLQMRNRNDV 933
 544 IQSESP-EVTSFPEPDVDSLMTTPPLPVVAGRPRLKLPQ-NFELPMID 593
 934 LVALDPLRASTFYPLPEDIAIOFVNEVTOAGENVNMEARDDEFGVPMVD 985

RESULT 2

NFH_HUMAN STANDARD; PRT; 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Ref. 12, Created)
 DT 01-OCT-1989 (Ref. 12, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 DE (Neurofilament heavy polypeptide) (NF-H).
 GN NEFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=68328981; PubMed=1138108;
 RA Lees J.F., Shneiderman P.S., Skuntz S.F., Carden M.J., Iazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 RT subunit (NF-H) and the gene encoding it.";
 RL EMBO J. 7:1947-1955(1988).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X15306; CAA33366.1; -
 DR EMBL: X15307; CAA33366.1; JOINED.
 DR EMBL: X15308; CAA33366.1; JOINED.
 DR EMBL: X15309; CAA33366.1; JOINED.
 DR PIR: S00979; QFNUH.
 DR Genew: HGNC:7737; NEFH.
 DR MIM: 162230; -
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT DOMAIN 1 100 HEAD.
 FT DOMAIN 101 413 ROD.
 FT DOMAIN 414 1020 TAIL.
 FT DOMAIN 101 132 COIL 1A.
 FT DOMAIN 133 145 LINKER 1.
 FT DOMAIN 146 244 COIL 1B.
 FT DOMAIN 245 266 LINKER 12.
 FT DOMAIN 267 288 COIL 2A.
 FT DOMAIN 289 292 LINKER 2.
 FT DOMAIN 293 413 COIL 2B.
 SQ SEQUENCE 1020 AA; 111780 MM; 1177G9DCB3DCE1D4 CRC64;

Query Match 6.18; Score 199; DB 1; Length 1020;
 Best Local Similarity 23.68; Pred. No. 0.04;
 Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

90 GQOEESSGGSVPLCPPPATKQAGIGGEPAAAGAGSPRRKYQAVPI-OTGSLVAAKE 148
 489 GEEDEAGGEEETKSP-----AEBAAS-----PEKEKSPVKEAKSPAEKAS 532
 149 PTPWAGDKGGAASPAAT---ASDPA---GPPPLPLPGPPPLATATAGTLAASGR 198
 533 P-----EKEAKSPAEVKSPEKAKSPEAKSPEAKSPEAKSPAEVKSPEAKSP 587
 199 WKMKRSPPLGGGSGGASQAACIKQILLQDLIEQOQOQLOAKKEIEBELKSERDTL 258
 588 AKKEAKSPAEKSPAEKSPVK-----EAKSPAEKSPVKEAKSP----- 629
 259 ARIERMRMQVLVKKDKNERHKLFOGTEEREETELSKIKLECOPELSETSQTLPPK 318
 630 AEVKSPPK-----AKSTYKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPV 664
 319 PFCGSRGSGKHKRSPGSTERKTPYKLAPEFSKYKTRPKHSPIKEE-PCGSLSETVC 377
 685 K---AEAKSPAEKSPV-KEAKSPAEKSPVKEAKSPAEKSPVKEAKSPVKEAKSPV 740
 378 KRELRSQ---TEKPRSSVDTPPRUSTPOKSGSTHP---KKAKSSLEEDLPYSTTE 430
 741 KEAKSPAEKSPAEKSPAEKSPVKEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPV 800
 431 MYLCRMHQPPSPPLURE---SSPKKE---TVARCLMPSSVAGETSVLAIVPSW---RDHS 482
 801 SPLKADAKAEKELPKKEEVKSYKKEEKQEVKVEPKKAEKAPATPKTEKKDKS 860
 483 VE---PLRDPNPSSDLENLDVSFKRAKLELDEKRRMDIORI---REQRILOQLRM 538
 861 KEAPKPEAKPKPYEKEKPEVPEKESKVE---AKKEAEDKKKVPTEKAPAKVEV-- 916
 539 YKKKGIOSEPEVTSFPEPDVDS 563

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 14, 2003, 11:47:07 ; Search time 24 Seconds

(without alignments)
1061.103 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238

Sequence: 1 MTRSAVFKAAPAGCNPE.....RSRCRLTIQKQTPHRTCRK 614

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211	6.5	1039	MSL1_DROME	P50535 drosophila
2	199	6.1	1020	NFH_HUMAN	P12036 homo sapien
3	197.5	6.1	831	NFH_RAT	P16884 rattus norv
4	197	6.1	1083	T2D3_HUMAN	O00268 homo sapien
5	196.5	6.1	2715	TRX2_HUMAN	O9um66 homo sapien
6	193.5	6.0	1723	ATM1_HUMAN	O9y4k1 homo sapien
7	192.5	5.9	520	WASP_MOUSE	P73315 mus musculu
8	190	5.9	1087	NFH_MOUSE	P19246 mus musculu
9	182.5	5.6	2161	SHK1_HUMAN	O9y566 homo sapien
10	181.5	5.6	1386	ZAP3_MOUSE	O9y017 mus musculu
11	178.5	5.5	2167	SHK1_RAT	O9y448 rattus norv
12	174	5.4	755	RRE1_HUMAN	O92766 homo sapien
13	173	5.3	351	CSP_PLAKU	P04922 plasmodium
14	172.5	5.3	707	SFRP0_HUMAN	P23246 homo sapien
15	172	5.3	3969	HRX_HUMAN	O03164 homo sapien
16	171	5.3	777	RN11_RAT	O64548 rattus norv
17	171	5.3	1647	SN24_HUMAN	P46634 homo sapien
18	169.5	5.2	616	VGE_HUMAN	O15240 homo sapien
19	169.5	5.2	857	NFM_CHICK	P16053 gallus gall
20	169	5.2	268	CDX1_MOUSE	P16111 mus musculu
21	168.5	5.2	1183	DRPL_RAT	P34258 rattus norv
22	168.5	5.2	3530	MY15_HUMAN	O9ukn7 homo sapien
23	167	5.2	2142	BAT2_HUMAN	O96634 homo sapien
24	166	5.1	2220	CAB1_HUMAN	O9y6j0 homo sapien
25	165	5.1	358	CEBA_RAT	P05554 rattus norv
26	165	5.1	1362	BRD4_HUMAN	O06085 homo sapien
27	165	5.1	1509	GSRI_HUMAN	O9uzm4 homo sapien
28	164.5	5.1	757	TAU_HUMAN	P10636 homo sapien
29	164	5.1	806	MK07_MOUSE	O9yvs8 mus musculu
30	164	5.1	1618	NKST_HUMAN	P46681 homo sapien
31	163	5.0	502	WASP_HUMAN	P42768 homo sapien
32	163	5.0	1217	AR4_MOUSE	O88573 mus musculu
33	163	5.0	2472	NCR2_MOUSE	O9wu42 mus musculu

34	162.5	5.0	1300	SAL3_HUMAN	O9bxa9 homo sapien
35	162.5	5.0	1446	IE18_PRYKA	P33479 pseudorale
36	162	5.0	817	VRP1_YEAST	P37370 saccharomyc
37	161.5	5.0	758	HUNB_DROME	P05084 drosophila
38	161.5	5.0	1262	CA13_CHICK	P12105 gallus gall
39	161.5	5.0	1461	IE18_PRTIF	P11675 pseudorale
40	161	5.0	579	SOC7_MOUSE	O8vhw2 mus musculu
41	161	5.0	1394	CNG4_BOVIN	O28181 bos taurus
42	161	5.0	2182	CAB1_RAT	O88480 rattus norv
43	160.5	5.0	1185	DRPL_HUMAN	P54259 homo sapien
44	160	4.9	1130	AB11_HUMAN	P00519 homo sapien
45	159.5	4.9	359	CEBA_MOUSE	P35366 mus musculu

ALIGNMENTS

RESULT 1
ID MSL1_DROME STANDARD: PRT; 1039 AA.
AC P50535;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Male-specific lethal-1 protein.
MSL-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE OF 85-1039 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93314941, PubMed=8325488;
RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the X chromosome in
RT males";
RT Genetics 134:545-557(1993).
RL [2]
RM REVISIONS, SEQUENCE FROM N.A.
RP MEDLINE=95300219, PubMed=7781064;
RX Kelley R.L., Solovayeva I., Lyman L.M., Richman R., Solovay V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
RT regulators on the X chromosomes and female lethality in Drosophila";
RT Cell 81:867-877(1995).
RL
CC -!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND
CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
CC PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
CC BINDING.
CC -!- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
CC SITES IN MALES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L42514; AAA8918.1; -
CC Flybase: FBgn0005617; msl-1.
CC Nuclear protein.
CC CONFLICT 188 193 PUPPA -> HCHLIP (IN REF. 1).
FT

DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE High molecular-weight neurofilament.
GN NF-H.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91038277; PubMed=2230956;
RA Chin S.S., Liem R.K.K.;
RT "Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vimentin in a predominantly nonphosphorylated form.";
RL J. Neurosci. 10:3714-3726(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Chin S.S.M., Liem R.K.K.H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL: AF031879; AAB87068.1; -
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
DR Coiled coil; Intermediate filament.
SQ SEQUENCE 1072 AA; 115349 MW; 89A146D457AD78B CRC64;

Query Match 6.1%; Score 197.5; DB 11; Length 1072;
Best Local Similarity 22.1%; Pred. No. 0.00082;
Matches 112; Conservative 82; Mismatches 235; Indels 77; Gaps 20;

QY 11 AAAPAGGNPQRLDYERAAALGPEDEPGAEEHFLPRHRRK---LKEPGPIASSGGS 66
DB 548 AEAKSPAEAKSPAEVKSPTATVKSPEAKSPAEAK-SPAELVKSPTATVKSPEAKSPAEAKS 606
QY 67 PAPSAGCGGKGRGLLPAGA-----APGQEESSGVSPLPCPPATKQAGTIGE-P 118
DB 607 PAEVKSPVEAKS-----PAEAKSPASVKSPEAKSPAEAKSPAEVKSPTATVKSPEAKSP 661
QY 119 AAAGAGCSPRKYQAVLPIDTGSILVAANKPEPTPWAGKGAASPAATAADSPAGPPLPLP 178
DB 662 AEVKSPTVKSPEAKSPVEVKS-PASVKSPE-ASPPAGAKSPAEAKSPVAKSPAEAK 719
QY 179 GPPPLPATAAGTLAASEGKWSKSPGLGGGSGSAOACIKQLLLQDLIEGQQQ 238
DB 720 SPAEAKPPAEKSPAEAK-----KSPAFAKSPAEAKSPAEKSPV---EVKSPAEAKS 768
QY 239 QLOAKKEIELEKSERDTLLARIERMRMOLYKKNDEKRRHKLFGQYETEEREETLSE 298
DB 769 PVKEGASLSAEKSPKAK-----KSPVKEIEIKPPAEVKSPEAKSPKME 812
QY 299 KIRLECPLESETSQTLPPKPFSGSGRSGKHRRKSPGSTER-KTPVKTLAPESKSYKTK 357
DB 813 EAK-----SPEAKTLDVKSPEAKTPAK--EBAKRPADIRSPQYKSPAEKSPKPEKETR 867
QY 358 TPKHSPITKEPSCSLSETVCK-----RELRSQETPEKPSV-----DTPRLSTP---QK 405
DB 868 TEKVAPEKKEVKSVEEVKAKKPPKVEEKTPATPEVKEKSKDAPAEAKPKAEK 927
QY 406 GPSTH-PKEKAFSEIEDLPYLSTTEKYLCRMHOPPPSPPLR-ESSPKKEEYVARCLMP 463
DB 928 EPLTEKPKDSPGEAKKEE-----AKEKKAAPAEETEPKALGVKEEAKPKAEKADAKAEK 982
QY 464 SVA-----GETSLAVPSWRDHSVE 484
DB 983 SKSESEKPKKEVPAPEKKDKTEE 1008

RESULT 14
ID 0921R1 PRELIMINARY; PRT; 2157 AA.
AC 0921R1;
OX 0921R1;

DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE BAT2.
GN NFKB1L.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
RA Shaffer T., Ratcliffe A., Loretz C., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility class III region.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF109719; AAC82480.1; -
DR MGD; MGI:1340031; Nfkb1l1.
SQ SEQUENCE 2157 AA; 229072 MW; 35B21F6B7C91F40 CRC64;

Query Match 6.1%; Score 197.5; DB 11; Length 2157;
Best Local Similarity 23.5%; Pred. No. 0.0017;
Matches 124; Conservative 49; Mismatches 213; Indels 141; Gaps 25;

QY 17 GNPEQRLDYERAA-----ALGG--PEDEPGAEEHFLPRHRRKLEP-----GPP 59
DB 496 GADPKRLKAPPAAPVTPAPALPVPVKEIPAPALPPTPTPEKEPEEPAQAPVQA 555
QY 60 ASSQGGSPAP---SPAGCGGKGRGL-----LLPAAAGGQEESSGGSVPLPCPP 106
DB 556 APSPGAVPPTLVLSGGGCTANSSSGSFASPVPEPOLISKEGPEPPE-----VPPETTP 610
QY 107 PATKQA-----GIGG--EPAAAGAGCSPRKYQAVLP-----IOTGSILVAANKPEPTWAG 155
DB 611 PAPKMEKGGGVSTRQPPSGIG--YPRYKQSLPFRPFRQGGQDLKQGGQGGQGGQ 667
QY 156 KGAASPA-----ATASDPAGPPLP---LPG---PPLAPATAGTLAASE 196
DB 668 QQGAPPAAPVPPSPQPVTLGAVPAPAPAPPPKALPGALGRPPMP-----MNF 720
QY 197 GRMKSM-----RKSLG-----GGGSGSSQAACIKQLLLQDLIEGQQQLOA 242
DB 721 PRMMMPPIPYDPLRLOGRPPLDTPPGVHSG-----LVPRESDSGSSSEP 768
QY 243 KEREIELEKSERDTLLARIERMRMOLYKKNDEKRRHKLFGQYETEEREETE-LSERIK 301
DB 769 FERRAPPLRERG-----PPVDPKLAW-----GDVFTTTPDPRRLTSLRQAD 815
QY 302 LECQPELSETSQTLPPKPFSGSGRSGKHRRKSPGSTERKTPVKKLAP----- 349
DB 816 EEEKSMSETPPVPPPPPYLA--NYPGFENGTPGPPIISFPLEESAPPGPRPLPWP 873
QY 350 -EFSKYTKTPKSPISKEEPCGSLSETVCKRELRSQETPEKPSVDTPTPRLSTPQKPS 408
DB 874 DEAAKMAOPPKPKPEKSEEPOLSGPAGKPPARGGGPPPPRRNRTETRW-----GPR 928
QY 409 THPKKAFSEIEDLPYLSTTEMYLCRMHOPPPSPPLRRESSPKKE 455
DB 929 PGSCRGRIPEEPGVPPRAGPI-----KKPPPVYVELLPKSLSE 969

RESULT 15
ID 088737 PRELIMINARY; PRT; 3942 AA.
AC 088737;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE Bassoon.
GN BSN OR BASSOON.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;


```

Db      809 VSEPTSTTANLLGLLSRSPVCPRLPGPENHRAGSNEDPILASGTPPTIPDETIF 868
QY      127 ----PRPKY-----QAVLPIONG-----LVAAKEPPWADKGGAAPATASD 168
Db      869 GGVNPPAVHYHKEASADVEISLESDDSDSVYVEBGLPPLPPPPSGATPPPIAPTGP 928
QY      169 PGPPLPL-----PGPPLPATATAGTLAASDGRKSMKSPGLGGGS 213
Db      929 PTASPPVPAKKEPEELPAAGPLPPEPPPPVPGVTLPPPG-----LVPEGTPCGG 981
QY      214 GASQAACLKQILLDLDIEQQOQLOAKKEIEELKSRDTLLARIEMRRMQLVKD 273
Db      982 GPALBEDLTVININSDEEEEGEEEEEDEEEEEE-----EEEDDFEEEE 1036
QY      274 DNE---KERHKLFGYETE-EREETESEKIKLECOPELSETOTLPKFGSGRGKCH 329
Db      1037 DEEEYEEEEEEEEEEEEEDEEGELEEEEEEDEEEEEEV---EDLEPTAG--- 1090
QY      330 KRKSPGSTERKTPVKLAPEFSKVTKTPKHSPIKEEPCGSLSETVCKRELKSOETPEK 389
Db      1091 -----GEVEGAPPPPTLPALPPESPCKVQPEPEPEGLLEV-----EEGTE 1136
QY      390 PRSSVTPPRLS-----TPQKGPSTHPKEKASSETEDLPYLSTTEMYLC 434
Db      1137 EERGAOTAPTLAEALPSOGEEVEREGESPAAGP---PPOELVEEESXPPPTLLEETEDG 1193
QY      435 RMHQPSPPLPLRESSPKKEETVARCLMPSVAGETSVLAVPSMRHSVEPLRDPSPD 493
Db      1194 SDKVQPPPTPAEEF--METETAEALOEKEQDDTAAMLA--DFIDCPPDEKPPPTPE 1248

```

RESULT 10

Q9UJ57 PRELIMINARY; PRT; 1026 AA.

```

ID 09UJ57 09UJ57; 09UJ57; 13, Created)
AC 09UJ57 09UJ57; 09UJ57; 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE Neurofilament protein (K1A0845 protein).
GN K1A0845
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhu Y., Han Y.
RT "Molecular Cloning of human hSTE cDNA."
RL Pei-Ching I Ko Ta Hsueh Hsueh Pao 31:531-531(1999).
RN [2]
RP SEQUENCE OF 94-1026 FROM N.A.
RC TISSUE-BRAIN:
RX MEDLINE-99156230; PubMed-10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka K., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 5:355-364(1998).
CC - SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL; AF203032; AAF13722.1;
DR EMBL; AB020652; BAA74868.1;
DR InterPro; IPR001386; Histone_H1/H5.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR00624; HISTONEHS.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament.
SQ SEQUENCE 1026 AA; 112479 MW; 0879B6A08D208C17 CRC64;

```

Query Match 6.28; Score 200; DB 4; Length 1026;
Best Local Similarity 22.48; Prid. No. 0.00055;

```

Matches 116; Conservative 69; Mismatches 228; Indels 104; Gaps 21;
QY      90 GOOESMGSVLPCCPPATKQAGIGCEPAAGAGCSPPKQOAVPI-OTGSLVAAKE 148
Db      489 GEEELGEGEETKSP-----AEKAS-----PEKAKSPVKEAKSPKEAKS 532
QY      149 PTPWADKGAASPAAT-----ASDPA-----GPPPLPLPGPPLPATATAGTLAASGR 198
Db      533 P-----EKEDAKSPAVKSPPEKAKSPAEKASPEKAKSPAEKASPEKAKSP 587
QY      199 WKMSKSPGLGGGSSAQAACLKQILLDLDIEQQOQLOAKKEIEELKSRDITLL 258
Db      588 AKKASPAEAKSPPEKAKSPV-----EAKSPAEKASPVKEAKSP----- 629
QY      259 ARTERERRMQLVKD-NEKERHKLFGYETEEREETESEKIKLECOPELSETOTLP 317
Db      630 AEYKSPKAKSPKEAKSPPEKAKSPPEKAKSPPEKAKSPPEKAKSPVKAASPEAKSP 689
QY      318 KPSCGRSGKHKRSPGSTERKTPVKLAPEFSKVTKTPKHSPIKEE-PCGSLSETV 376
Db      690 VK---AEKSPPEKAKSPV-KEAKSPPEKAKSPVKEAKSPPEKAKSPVKEAKSP 745
QY      377 CKRELKSOE---TPKPRSSVTPPRLSTPQKGPSTHP---KEKAFSEIEDLPYLSTT 429
Db      746 VKKASPEKAKSPPEKAKSPPEKAKSPPEKAKSPPEKAKSPPEKAKSPVKEAKSP 799
QY      430 EMYLCRMHQPSPPLPLRESS-----PKKEETVARCLMPSVAGETSVLAVPSMRDS 482
Db      800 -----KSPKAKSPKEDAKPEKEITKKEE-VASPVKEEKPEVVKPEKPAEE 851
QY      483 VEPLRDPNPSDLENDLSVFSRRHAKLEDEKRRKRWMDQIRLEQRIQLRLQRLMYKK 542
Db      852 KAP-ATPKTEKKKSKKEAPKKEAPKPEVEEKKEPA--VEKKEKSVAKKEEADKK 908
QY      543 GIDSEPEVTSF-----PEPDVVS 563
Db      909 -VPTPEKAPKAVKEDAKPKKEKTEVAKKEDDADA 944

```

RESULT 11

Q9R0R5 PRELIMINARY; PRT; 1034 AA.

```

ID 09R0R5 09R0R5; 09R0R5; 13, Created)
AC 09R0R5 09R0R5; 09R0R5; 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE Transcription factor CA150b.
GN TCERG1 OR TAF2S.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimada M., Horigome T.;
RC TISSUE-LIVER:
RX MEDLINE-20047056; PubMed-10578054;
RA Shimada M., Saito M., Katakai T., Shimizu A., Ichimura T., Omata S.,
RA Horigome T.;
RT "Molecular cloning and splicing isoforms of mouse p144, a homologue of
RT CA150."
RL J. Biochem. 126:1033-1042(1999).
DR EMBL; AB023485; BAA86392.1;
DR MGD; MGI:1926421; Tcerg1.
DR InterPro; IPR002713; FF.
DR InterPro; IPR002965; P-rich_extensions.
DR InterPro; IPR001202; WW_Rsp5_WWP.
PFam; PF01846; FF; 5.
DR PFam; PF00397; WW; 3.

```

Query Match 6.3%; Score 203.5; DB 4; Length 2432;
 Best Local Similarity 22.7%; Pred. No. 0.00087;
 Matches 133; Conservative 79; Mismatches 254; Indels 121; Gaps 23;

QY 6 AVFAAANA---GANEQRLDYRAAALGPEDEPGAAEHFLPRHKLKEPGLASS 62
 DB 209 AVSAGAPASQCAQGPQOV-----PVTPQAPCPQHAPPHQSKAQ-----AEM 254
 QY 63 QGSGPAP-----SPAGCGGKRGILLPAGAAPGQOEGSGSVPL--PCPPATQAGTGG 116
 DB 255 QSRLEPGTARYSAAGSEAKTRGSGTTANNRSQSFNNDKSKPVTSPPPSHKK----- 310
 QY 117 EPAAGAGCSPRPYQAVLPQTGSLVAAAKEPTWAGDKGAA-----SPAATASDPAG 171
 DB 311 EPLASASSHPGMDNAPASLESGSSSPPTCSTSSAIPQGAATKPPRSLSVKHSAT 370
 QY 172 PPPLPLPPPLPAPTAAGTLAASEGRKMSMK-----SPLGGGGSGASSQAACLKQI 225
 DB 371 VSMLSVPPGPEAPRPPEAKMPAPNNQSMLEKTLFNSKGSKAGGPGSRDSCERL 430
 QY 226 LLQLDLLEQOQOOLAKKEIEEL-----KSERDTLARIEMRRMQLVKD-----NEK 277
 DB 431 ETLF-----SESESELEAASRLTTPVGPASSPKTALAGIAQRTSRALTNKSSLKGNK 487
 QY 278 ERHKLFGQYETEEEEE-----TELSEKIKLECOPELSETQTLPPKPFSCGRSG---KGH 329
 DB 488 EKEQOQREKDEKDLAKRASVTERLDLKEPEDEPGAAVPEMPKSSKIASFIRPGG 547
 QY 330 K-----KSPFGSTERTKTPVKLAPEFSKVTKTPK--HSPKEEPCGSLSETVCKRELRSQ 384
 DB 548 KLSAKKEPMAPSHSGIP---KPGKMSMPGKSPSAPAPSK-----GERSRSGKLSG 597
 QY 385 ETPKRP-----SSVDTPPRLSTPOKGPSTHKEKAFSE-----IEDLYLSTTEKYL 433
 DB 598 LPQKPPOLDGHHSSSSSLASSEKPGCGTTLNHSISQTVSGSVGTQTGTSMTVSQVL 657
 QY 434 CRWQPPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSMRDSVEPL----- 486
 DB 658 -----PQPOQYNNPNTAVAPFLYRSQDTEGVATPSSSTGVSVEPSHPTKQG 708
 QY 487 -----RDP-----NPSDLNLDSDVFSKR-----HAKLE 511
 DB 709 PALEELTGEDPEARRLRTVKNIADLRQMLEETMSLSRGTQYHSTLE 755

RESULT 8

ID 015451 PRELIMINARY; PRT; 1021 AA.
 AC 015451;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Proline and glutamic acid rich nuclear protein isoform (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vadlamudi R.K., Chung J.K., Shin J.;
 RT "p160.2 an isoform of proline and glutamic acid rich nuclear protein p160";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U88154; AAC17709.1; -;
 DR InterPro; IPR002965; P_rich_extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 KW Nuclear protein.
 FT NON_TER 1
 SO SEQUENCE 1021 AA; 109136 MW; C895184BD67698D3 CRC64;

Query Match 6.2%; Score 200.5; DB 4; Length 1021;

Best Local Similarity 22.1%; Pred. No. 0.00051;
 Matches 119; Conservative 51; Mismatches 236; Indels 133; Gaps 20;

QY 38 PGAAEAHFLPRHRK-----LKEPGPLASSQGGSPAPSPAGGKRGILLPAGAPG 90
 DB 530 PTPAPVPLLRHPSGPHRSITLRAPCPQMA-----PCPOQAPCPASAG-----PMSAGP 578
 QY 91 QOESWGG-----SVPLPCPP-----PATQAGIGGPAAGAGCS----- 126
 DB 579 VPSEPWTSTANLGLSRPVCPRLLPGENHRASGNEDPLAPSGTPTPTTPPETE 638
 QY 127 ----PRKY-----QAVLPQTGS-----LVAAKEPTWAGDKGAAAPASD 168
 DB 639 GGRVPRAPFAVHYDEKASDVEISLSDSDSVIVPEGLPLPPPPSPGAPPTPIAPTGP 698
 QY 169 PAGPPPLPL-----PCPPPLAPPTAGTLAASEGRKMSMKRSPIGGGGS 213
 DB 699 PTAAPPVPAKEPEELPAPGPLPPPPPPPPVPGVTLPPQ-----LVPGTGGG 751
 QY 214 GASSQAACLKQILLQLDLLEQOQOOLAKKEIEELKSERDTLARIEMRRMQLVK 273
 DB 752 GPPLAEDLYININSSDEEEEGEGEBEEEBEEEBEE-----EEEDDFEEDE 806
 QY 274 DNE---KERHKLFGQYETE--EREETESEKIKLECOPELSETQTLPPKPFSCGRSGKH 329
 DB 807 DEEYFEEEBEEEBEEEBEEEBEEEBEEEBEEEBEEEBEE---EDLEFGTAG--- 860
 QY 330 KRKSPFGSTERTKTPVKLAPEFSKVTKTPKHSPIKEEPCGSLSETVCKRELRSQETPEK 389
 DB 861 -----GEVEGAPPPTPLPALPPSPPPVQPEPEPEPLLEV-----EEGQTE 906
 QY 390 PRSSVDTPPRLS-----TPQKGPSTHKEKAFSEIEDLYLSTTEKYL 434
 DB 907 EERADDTAPTLAPALPSQGEVEREGESPPAGP---PQGLVEEESPXPRTLLLEETEDG 963
 QY 435 RWHQPPSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSMRDSVEPLRDPNSD 493
 DB 964 SDKVQPPPEPTAEAE--METETEAALQEKQDQDTAAMLA--DFIDCPDDEKPPPPTE 1018

RESULT 9

ID 015450 PRELIMINARY; PRT; 1251 AA.
 AC 015450;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE p160 (Fragment).
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vadlamudi R.K., Chung J.K., Shin J.;
 RT "Cloning and expression of a cDNA for p160 a nuclear, proline rich protein that interacts with Ick SH2 domain";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U88153; AAC17708.1; -;
 DR InterPro; IPR002965; P_rich_extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 FT NON_TER 1
 SO SEQUENCE 1251 AA; 134100 MW; 21E0F8549FE03011 CRC64;

Query Match 6.2%; Score 200.5; DB 4; Length 1251;
 Best Local Similarity 22.1%; Pred. No. 0.00064;
 Matches 119; Conservative 51; Mismatches 236; Indels 133; Gaps 20;

QY 38 PGAAEAHFLPRHRK-----LKEPGPLASSQGGSPAPSPAGGKRGILLPAGAPG 90
 DB 760 PTPAPVPLLRHPSGPHRSITLRAPCPQMA-----PCPOQAPCPASAG-----PMSAGP 808
 QY 91 QOESWGG-----SVPLPCPP-----PATQAGIGGPAAGAGCS----- 126

DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Unknown (Protein for IMAGE:4156419) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC010457; AAL10457.1; -
 FT NON_TER
 SQ SEQUENCE 1052 AA; 112161 MW; C71A751164C8C1AB CRC64;
 Query Match 6.5%; Score 209.5; DB 4; Length 1052;
 Best Local Similarity 23.1%; Pred. No. 0.00015;
 Matches 115; Conservative 51; Mismatches 204; Indels 127; Gaps 20;
 QY 55 PGPPPLASOGG-----SPAPSPAGCGGKRGGLLPAGAAPG--QOESWGS 99
 Db 622 PGPPPTANHLGLSVPLVSVPRLLPGPENHRAGSNEDPLASGTPPTIPDETFTGR 681
 QY 100 VPLPCPPPATKQ-----AGIGEPMAAGACSPRPKYQAVLP10 138
 Db 682 VPRPAVHYDKDEASDVEISLESDDSVYIVPEGLPPLPPPPSGATPPP----- 732
 QY 139 TGLVAAAKPEPTWADKGAAPATASD-----FAGPPPLPGPPPLAATAGT1A 193
 Db 733 ----LAPGPPPT-----ASPPVPAKEEPBELPAAPGLP-PPPPPPPPGPPVTL 778
 QY 194 ASEGKMSKMSKSLPGGSGGSSQAACLKQILLDLLEQOOOLQAKETIEELKSE 253
 Db 779 PPG-----LVPEGTGGGGPPALEEDLVININSSD--EEEEEEEEEEEEEEEE 829
 QY 254 RDTLLARIERMERMOLVKKDKNEKRLKLFQYETEERETELSEKIKLECOPELSETSQ 313
 Db 830 ED-----FEEEDDEEYEEFEEDDEEFEEDDEEFEEDDEEFEEDDEEELBEV-- 882
 QY 314 TLPPKPFSGSGKSGKSRSPFSTERKTPVKTLAEPFSVKTKTKHSIKKEPGCSLS 373
 Db 883 ----EDLEFETAG-----GEVEGAPPPPTLPALPPSPPKVQPEPEPEGLLL 929
 QY 374 E-----TWCKRELRSQ-ETPEKPRSSVDTPRLSTPQKPSYHPEKAF 416
 Db 930 EVEPEPTGEERGADTAPTLAPALPQGEVEREGESPAAGPPQELVEEPSAP--TLL 987
 QY 417 SSEIEDLPYLSTEMTLCRMHOPPSPLPRESSPKKEETVANCLMPSSVAGETSVLAVP 476
 Db 988 EEEETED-----GSDKV-----QPPPE-TPAEEE--METETEALQKEQDDTAAMLA-- 1032
 QY 477 SWRDHSEVPLRDPNPSD 493
 Db 1033 DFTDCPPDDEKPPPPTE 1049
 RESULT 6
 OBTDF0 PRELIMINARY; PRT; 2429 AA.
 AC OBTDF0;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Retinoic acid inducible in neuroblastoma cells RAIN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=21902045; PubMed=11904404;
 RX MEDLINE=21902045; PubMed=11904404;

RA Merrill R.A., Plum L.A., Kaiser M.E., Clagett-Dame M.;
 RT "A mammalian homolog of unc-53 is regulated by all-trans retinoic acid
 RT in neuroblastoma cells and embryos."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:3422-3427(2002).
 DR EMBL: AF466143; AAL96479.1; -
 SQ SEQUENCE 2432 AA; 261380 MW; DAC37362819DE2DF CRC64;
 Query Match 6.3%; Score 203.5; DB 4; Length 2429;
 Best Local Similarity 22.7%; Pred. No. 0.00087;
 Matches 133; Conservative 79; Mismatches 254; Indels 121; Gaps 23;
 QY 6 AVFKAAAPR---GNPEQRLDYERAAALGPEDEFGAAMFLPHRRLKEPGPLASS 62
 Db 209 AVSQVAGAPSQCAQAPQOQV-----PVTPOAPCQHPAPAHQOSKAQ---AEM 254
 QY 63 QGGSPAP---SPAGCGGKRGGLLPAGAAPGQOESWGSVPL--PCPPPATKQAGIG 116
 Db 255 QSRPGPTARVASAGSEATRTGGSTTANNRRSSFNNDKSKPVTSPPPSSHEK----- 310
 QY 117 EPAAGAGCSPPRPKYQAVLP10TQSLVAAAKEPTPWADKGAAPG-----SPATASDPAG 171
 Db 311 EPLASSASSHPGKSDNAPASLESSTFTNCSTSSAIPQGAATKPMRSKLSVYKHSAT 370
 QY 172 PPLPLPGPPPLAATAGT1AASBGRKMSMK-----SPUGGGGSGAQAACLKQ1 225
 Db 371 VSMISVYKPPGPEAPRPTPEAMKPAANNOKSMLEKTLFNSKGSKAGEGPGSRDTSCERL 430
 QY 226 LILQDLILGQOOOLQAKKEIEEL-----KSEBDTLARIERMERMOLVKD-----NKR 277
 Db 431 ETLP---SFESEBELASMLTTPVGPASSPPIAKLGNQRFSSALTNKSKSLKGNK 487
 QY 278 ERHKLQGYETEREE---TELSEKIKLECOPELSETSQTLPPKPFSGRSG---KQH 329
 Db 488 EKQKQKREKDKSKDLAKRASVTERLDLKEEPKEDPSGAAYEMPKSKSIASFIPKG 547
 QY 330 K-----KSPGSGTERKTPVKTLAPFSKYKTKTPK--HSPKPEPCSLSEYVCKRLRQ 384
 Db 548 KLSAKKEPAPSHSGIP---KPGKSMGKSPAPAPSK-----GERSHSGLSGG 597
 QY 385 ETEKRP-----RSSVDPPRLSTPQKPSYHPEKAFSS-----IEDLPYLSTEMTL 433
 Db 598 LPQKQQLDGRHSSSSSLASSGKPGGTLNHSISQTVSGSVTTQTTGSGNTVSYQL 657
 QY 434 CRWHOPPSPLPRESSPKKEETVANCLMPSSVAGETSVLAVSWDHSVEPL----- 486
 Db 658 -----POPOQVNHNPATVAFVLRSDTDEGNTAESSGVSVEPSHFTKGTQ 708
 QY 487 -----NDP-----NPSDLLENLDDSVFSKR-----HAKLE 511
 Db 709 PALEELTGEDPEARLRLTVKNLADLRQNLLETWSSLRGTQVTHSTLE 755
 RESULT 7
 OBTDF0 PRELIMINARY; PRT; 2432 AA.
 AC OBTDF0;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Retinoic acid inducible in neuroblastoma cells RAIN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=21902045; PubMed=11904404;
 RX Merrill R.A., Plum L.A., Kaiser M.E., Clagett-Dame M.;
 RT "A mammalian homolog of unc-53 is regulated by all-trans retinoic acid
 RT in neuroblastoma cells and embryos."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:3422-3427(2002).
 DR EMBL: AF466144; AAL96480.1; -
 SQ SEQUENCE 2432 AA; 261722 MW; 2666957ED0A726F9 CRC64;

SQ SEQUENCE 616 AA: 67332 MW: 221C84C306C7D07D CRC64;
 Query Match 96.0%; Score 3110; DB 11; Length 616;
 Best Local Similarity 95.9%; Pred. No. 5e-177;
 Matches 591; Conservative 10; Mismatches 13; Indels 2; Gaps 2;

QY 1 MTRSAVFKAAAPAGNPNBQRLDYERAAALGPEDEFGAAEHFLPRHKKLEKPPPLA 60
 DB 1 MTRSAVFKAAAPAGNPNBQRLDYERAAALGPEDEFGAAEHFLPRHKKLEKPPPLA 60
 QY 61 SSGGSPAPSPAC-GGKGRGLLPAGAAAGQEESSGVPLPCPPATKAGIGCEPA 119
 DB 61 SSGGSPAPSPACGGKGRGLLPAGAAAGQEESSGVPLPCPPATKAGIGCEPA 120
 QY 120 AAGAGCSPRKYQAVLPIDTGSITVAAPAKETPMAGKGAAPPAATASDPAGPPPLP 178
 DB 121 AAGAGCSPRKYQAVLPIDTGSITVAAPAKETPMAGKGAAPPAATASDPAGPPPLP 180
 QY 179 GPPPLATATAGTAAAGSGRMSKMSPLGGGGSGASSQAACILQILLQLDLIEQOOQ 238
 DB 181 GPPPLATATAGTAAAGSGRMSKMSPLGGGGSGASSQAACILQILLQLDLIEQOOQ 240
 QY 239 QLOAKEKEIEELSERDTLLARIERMERMOVLKDKNEKRHKLFQGYETEEPEPELSE 298
 DB 241 QLOAKEKEIEELSERDTLLARIERMERMOVLKDKNEKRHKLFQGYETEEPEPELSE 300
 QY 299 KILCEQPELSETSQLPPKPFSGSGRSGKHKRPSTERTKTPVKLAPEFSKVTXT 358
 DB 301 KILCEQPELSETSQLPPKPFSGSGRSGKHKRPSTERTKTPVKLAPEFSKVTXT 360
 QY 359 PKHSPIKEEPCGSLSETVCKRELRSOETPEKPNSSVDTPPRLSTPOKGPSTHPEKAFSS 418
 DB 361 PKHSPIKEEPCGSLSETVCKRELRSOETPEKPNSSVDTPPRLSTPOKGPSTHPEKAFSS 420
 QY 419 EIEDLPPLSTTEWLCWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSW 478
 DB 421 EIEDLPPLSTTEWLCWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSW 480
 QY 479 RDSHVEPLRPNPSDLLENDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLM 538
 DB 481 RDSHVEPLRPNPSDLLENDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLM 540
 QY 539 YKKKGIOSEPEVTSFEPEDDVESLITPFLPVVAFGRPLKLTPONFELPWLDERSRC 598
 DB 541 YKKKGIOSEPEVTSFEPEDDVESLITPFLPVVAFGRPLKLTPONFELPWLDERSRC 600
 QY 599 RLEIÖKKOTPHRTCK 614
 DB 601 RLEIÖKKOTPHRTCK 616

RESULT 2
 Q9DSC9 PRELIMINARY; PRT; 370 AA.
 ID 09DSC9
 AC 09DSC9
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 4930463F05R1K protein.
 GN 4930463F05R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Srinivasan L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seyg T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK015496; BAB29688.1; -;
 DR MGI:1922153; 4930463F05R1K.
 SQ SEQUENCE 370 AA: 43184 MW: 2347FA88F39B5654 CRC64;

Query Match 52.8%; Score 1711; DB 11; Length 370;
 Best Local Similarity 89.3%; Pred. No. 3.4e-94;
 Matches 324; Conservative 10; Mismatches 13; Indels 16; Gaps 1;

QY 252 SERDTLLARIERMERMOVLKDKNEKRHKLFQGYETEEPEPELSEKILCEQPELSET 311
 DB 24 TNRKRLARIERMERMOVLKDKNEKRHKLFQGYETEEPEPELSEKILCEQPELSET 83
 QY 312 SQTLPKPFSGSGRSGKHKRPSTERTKTPVKLAPEFSKVTXTPKHSPIKEEPCGS 371
 DB 84 SQTLPKPFSGSGRSGKHKRPSTERTKTPVKLAPEFSKVTXTPKHSPIKEEPCGS 143
 QY 372 LSETVCKRELRSOETPEKPNSSVDTPPRLSTPOKGPSTHPEKAFSSIEDLPYSTTEM 431
 DB 144 LSETVCKRELRSOETPEKPNSSVDTPPRLSTPOKGPSTHPEKAFSSIEDLPYSTTEM 203
 QY 432 YLCWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSWRDSHVEPLRPNP 491
 DB 204 YLCWHQPPSPPLRLRESSPKKEETVARCLMPSSVAGETSVLAAPSWRDSHVEPLRPNP 247
 QY 492 SDLENDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLMYKKKGIOSEPEV 551
 DB 248 SDLENDDSVFSKRAKLELDEKRRKRWIORIREORIOLRLMYKKKGIOSEPEV 307
 QY 552 TSFEPEDDVESLITPFLPVVAFGRPLKLTPONFELPWLDERSRCLEIÖKKOTPHRT 611
 DB 308 TSFEPEDDVESLITPFLPVVAFGRPLKLTPONFELPWLDERSRCLEIÖKKOTPHRT 367

RESULT 3
 Q9VJ66 PRELIMINARY; PRT; 1039 AA.
 ID 09VJ66
 AC 09VJ66
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE MSL-1 protein (Fragment).
 GN MSL-1 OR CG10385.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 14, 2003, 11:47:10 ; Search time 87 Seconds
(without alignments)
1454.172 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3338

Sequence: 1 MTRSAVFKAAPAGNPE.....RSRCLEIQKKQPHRCRK 614

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteus:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriophage:*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	96.0	616	11	Q9CXF9 mus musculus
2	1711	52.8	370	11	Q9DSC9 mus musculus
3	213	6.6	1039	5	Q9VJ66 drosophila
4	209.5	6.5	1052	4	Q9B060 homo sapien
5	209.5	6.5	1052	4	Q96FT1 homo sapien
6	203.5	6.3	2429	4	Q8TDF0 homo sapien
7	203.5	6.3	2432	4	Q8RDE9 homo sapien
8	200.5	6.2	1021	4	Q15451 homo sapien
9	200.5	6.2	1251	4	Q15450 homo sapien
10	200	6.2	1034	11	Q9U057 homo sapien
11	200	6.2	1034	11	Q9R0R5 mus musculus
12	200	6.2	3938	11	Q88778 mus musculus
13	197.5	6.1	1072	11	Q35482 mus musculus
14	197.5	6.1	1257	11	Q9Z1R1 mus musculus
15	197.5	6.1	3942	11	Q88737 mus musculus
16	196.5	6.1	821	4	Q96JW3 homo sapien

17	194.5	6.0	1123	11	Q9DBD5 mus musculus
18	193.5	6.0	520	11	Q61078 mus musculus
19	193.5	6.0	1173	11	Q9QXV3 mus musculus
20	193.5	6.0	1197	11	Q9QXV2 mus musculus
21	193	6.0	753	4	Q9H987 homo sapien
22	192.5	5.9	762	4	Q9C0H9 homo sapien
23	192.5	5.9	1475	5	Q9N4G4 homo sapien
24	192.5	5.9	5085	11	Q9JKS6 mus musculus
25	192	5.9	3534	12	Q9Z266 mus musculus
26	192	5.9	3536	5	Q9VZ30 mus musculus
27	191	5.9	540	4	Q96B87 mus musculus
28	190.5	5.9	917	4	Q9P1Y5 mus musculus
29	190	5.9	1060	3	Q9C259 mus musculus
30	190	5.9	1098	4	Q14776 mus musculus
31	189.5	5.9	599	6	Q9SK67 mus musculus
32	189.5	5.9	606	6	Q28687 mus musculus
33	189	5.8	1154	4	Q9P2P0 mus musculus
34	189	5.8	1431	4	Q8WXP9 mus musculus
35	189	5.8	5147	4	Q9Y6V0 mus musculus
36	189	5.8	554	4	Q9HBC2 mus musculus
37	188.5	5.8	675	11	Q91M9 mus musculus
38	188.5	5.8	1800	4	Q9H1C1 mus musculus
39	188	5.8	1862	4	Q9H1C2 mus musculus
40	188	5.8	2703	5	Q9VEG7 mus musculus
41	188	5.8	2715	5	Q61603 mus musculus
42	188	5.8	966	4	Q8WU64 mus musculus
43	187.5	5.8	1644	4	Q96JK5 mus musculus
44	187.5	5.8	1644	4	Q96JK5 mus musculus
45	187.5	5.8	2701	4	Q9Y520 mus musculus

ALIGNMENTS

RESULT 1

ID	Q9CXF9	PRELIMINARY:	PRT:	616 AA.
AC	Q9CXF9	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	4121402D02R1K protein.			
GN	4121402D02R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Scleroglossa; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kakuwaka T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batlov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Bakei J., Oikido T., Furuno M., Aono H., Baldarelli R., Bafsh G.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK014463; BAB29369.1; -			
DR	MGD; MGI:1921276; 4121402D02R1K.			


```
Db 123663 ATTCAGTTTTCAGCTGAATGCCCTGCATTCCTAATAACATAGTTG-----TTTCTTT 123716
QY 4147 AACAAAGTTTAAGCTAGTGTATATAATTAATAAA--AAATTGCTGTGCTGCTACTTCA 4203
   ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123717 AATAAAGTTTACACTAATCTTTGTAAATCAGCAAGCAATTCCTAATCTGTCCACTTCA 123776
QY 4204 GCTTGTGTTTATGCCATTTCATATATGTTGTGTGTGTAAATTCATTAACCTTTTGATACC 4263
   ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123777 ACTTGTGTTTATGCCCATTTCAATGTGTGTGTGTGCAATTCATACTTTTGACACCA 123836
QY 4264 ATTCGATGTGTAATAATGCTGTCTGTAAATATCTATATAAGAGTCAATGTAAAT 4323
   ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123837 ATTCGATATGAATAATGTCATTCCTGTGT-AGTATCCTACGAAAGAGTCAATGTGAGGT 123895
QY 4324 AAACATGTGCTGTTAATAAAAAAAAAAAAAA 4358
   ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123896 AATTTATGTGCTGCTAATTTTGAACAAACAAA 123930
```

Search completed: July 14, 2003, 16:44:42
Job time : 10978 secs

VERSION	KEYWORDS
AL513172.1	GI:12539823
HTG: HTGS_PHASE1; HTGS_CANCELLED.	
SOURCE	ORGANISM
human.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	AUTHORS
TITLE	JOURNAL
COMMENT	
<p>Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk</p> <p>----- Genome Center Center: Sanger Centre Web site: http://www.sanger.ac.uk Contact: humquerry@sanger.ac.uk</p> <p>----- Project Information Center project name: BA291H23</p> <p>----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: dye-terminator Big Dye; 100% of reads Consensus quality: 137540 bases at least Q40 Consensus quality: 139579 bases at least Q30 Consensus quality: 140684 bases at least Q20 Insert size: 141308; sum-of-contigs Insert size: 162784; 14.9% error; agarose-fp Quality coverage: 4.49x in Q20 bases; sum-of-contigs Quality coverage: 4.05x in Q20 bases; agarose-fp</p> <p>----- NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p> <p>* * 11819: contig of 11819 bp in length * 11820 11919: gap of 100 bp * 11920 21066: contig of 9147 bp in length * 21067 21166: gap of 100 bp * 21167 32577: contig of 11411 bp in length * 32578 32677: gap of 100 bp * 32678 35204: contig of 2527 bp in length * 35205 35304: gap of 100 bp * 35305 38616: contig of 3312 bp in length * 38617 38716: gap of 100 bp * 38717 48887: contig of 10171 bp in length * 48888 48987: gap of 100 bp * 48988 51811: contig of 2824 bp in length * 51812 51911: gap of 100 bp * 51912 54004: contig of 2093 bp in length * 54005 54104: gap of 100 bp * 54105 60083: contig of 5979 bp in length * 60084 60183: gap of 100 bp * 60184 66204: contig of 8021 bp in length * 66205 68304: gap of 100 bp * 68305 72836: contig of 4532 bp in length * 72837 72936: gap of 100 bp * 72937 88769: contig of 15833 bp in length * 88770 88869: gap of 100 bp * 88870 98455: contig of 9586 bp in length * 98456 98535: gap of 100 bp * 98536 102201: contig of 3646 bp in length * 102202 102301: gap of 100 bp * 102302 128671: contig of 26370 bp in length * 128672 128771: gap of 100 bp * 128772 131983: contig of 3212 bp in length * 131984 132083: gap of 100 bp * 132084 142908: contig of 10825 bp in length.</p> <p>Location/Qualifiers</p>	
FEATURES	

```

source
1. .142908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-291H23"
/clone_lib="RPC1-11.2"
1. .11819
/misc_feature
/note="assembly_fragment:00446
clone_end:SP6
vector_side:left"
misc_feature
11920. .21066
/note="assembly_fragment:01125
fragment_chain:1"
21167. .32577
/note="assembly_fragment:01044
fragment_chain:1"
32678. .35204
/note="assembly_fragment:01155
fragment_chain:1"
35305. .38616
/note="assembly_fragment:00076
fragment_chain:1"
38717. .48887
/note="assembly_fragment:01431
fragment_chain:2"
48988. .51811
/note="assembly_fragment:01037
fragment_chain:2"
51912. .54004
/note="assembly_fragment:01067
fragment_chain:2"
54105. .60083
/note="assembly_fragment:01281
fragment_chain:2"
60184. .68204
/note="assembly_fragment:00264
fragment_chain:3"
68305. .72836
/note="assembly_fragment:00682
fragment_chain:3"
72937. .88769
/note="assembly_fragment:00228"
88870. .98455
/note="assembly_fragment:00556"
98556. .102201
/note="assembly_fragment:00790"
102302. .128671
/note="assembly_fragment:01503"
128772. .131983
/note="assembly_fragment:01404
fragment_chain:4"
132084. .142908
/note="assembly_fragment:00926
fragment_chain:4
clone_end:r7
vector_side:right"
BASE COUNT 45765 a 25742 c 25056 g 44744 t 1601 others
ORIGIN
Query Match 5.4%; Score 235.4; DB 2; Length 142908;
Best Local Similarity 81.8%; Pred. No. 4.2e-34;
Matches 323; Conservative 0; Mismatches 61; Indels 11; Gaps 4;

3968 TTGTGTGAGAGAGAGCTGGGACAGTAGATTACCAAGTTCACAGTCCAAAATTACAGTGT 4027
|||||
123543 TTGTGTGAGAGAGAGAGCTGGGAGTAGATTACCAAGTTCACAGTCCAAAATTACAGTGT 123602
|||||

4028 GTTAGAGTGTGGGGGAGAAATTAGTCTTTTTCCTCATGGGATACACACCTGTG-A 4086
|||||
123603 GTCAAGAGTGTGGGGGAGAAATTATACCTTATTTTTCCTCAAAATGGGATATACACTGTGA 123662
|||||

4087 ATTCAATCTCAACTGAAGGCCCTGCAGTTCTCTCAAAACATAGTGTGTTTCTTTCTTT 4146
|||||

```

Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

FEATURES
source
Derived from dbEST (genbank accession R78466).
Location/Qualifiers
1..240
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="319.5 CR from top of Chr17 linkage group"

STS
primer_bind 1..150
primer_bind 1..23
complement(127..150)

BASE COUNT 76 a 45 c 51 g 68 t
ORIGIN

Query Match 5.5%; Score 240; DB 11; Length 240;
Best Local Similarity 100.0%; Pred. No. 2.2e-35;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2943 AAAATATCAGATGACACCCACGACCTTGTCTCACTGTAAGATCCCTTTACTA 3002
DB 1 AAAATATCAGATGACACCTCCAGGCCACTTGTCTCACTGTAAGATCCCTTTACTA 60
QY 3003 TCTGAAGAGAAATAGACCCAGACCTCTGTCTCAATATATAGAAATGCTTTCTT 3062
DB 61 TCTGAAGAGAAATAGACCCAGACCTCTGTCTCAATATATAGAAATGCTTTCTT 120
QY 3063 AGCTTCAGACATATGTGTGAAGAACAGTAGGGGTCTAATCTCTTGAAGAGTAGGGCT 3122
DB 121 AGCTTCAGACATATGTGTGAAGAACAGTAGGGGTCTAATCTCTTGAAGAGTAGGGCT 180
QY 3123 TTATATCTTAAGAGATATGTCTCCGATTTATAGCACTTTAAGAGAGAGCAAGGT 3182
DB 181 TTATATCTTAAGAGATATGTCTCCGATTTATAGCACTTTAAGAGAGAGCAAGGT 240

RESULT 14
AL354659 135060 bp DNA linear PRI 15-NOV-2001
LOCUS Human DNA sequence from clone RP11-157G15 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL354659 AC024007
VERSION AL354659.14 GI:16972840
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 135060)
Griffiths, C.
Direct Submission
Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:1390411.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submissions
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-157G15 is from the library RP01-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-157G15. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-157G15 is at 135060 in this sequence. The true left end of clone RP11-291H23 is at 34561 in this sequence. The true right end of clone RP11-361K17 is at 2000 in this sequence.

FEATURES
source
Location/Qualifiers
1..135060
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-157G15"
/clone_1b="RP01-11.1"

BASE COUNT 44915 a 24717 c 23719 g 41709 t
ORIGIN

Query Match 5.4%; Score 235.4; DB 9; Length 135060;
Best Local Similarity 81.8%; Pred. No. 4.2e-34;
Matches 323; Conservative 0; Mismatches 61; Indels 11; Gaps 4;

QY 3968 TTTGTGAGAGAGAGGTGGACAGATAGCAAGTTCAGCAAAATTTACAGTGT 4027
DB 126046 TTTGTGAGAGAGAGAGATGGAAGTATAGTAATGTTCAAGAGTGAATACAGTGT 126105
QY 4028 GTTAAAGTGTGGGGGAAATTAATGTTATTTTCCCTACATGAGGATACAACTGTG-A 4086
DB 126106 GTTAAAGTGTGGGGGAAATTAATGTTATTTTCCCTACATGAGGATACAACTGTGAA 126165
QY 4087 ATTCATATCTCACTGAAAGCCCTGCAGATTCCTAAACATATGTTGTTTCTTT 4146
DB 126166 ATTCATATCTCACTGAAAGCCCTGCAGATTCCTAAACATATGTTGTTTCTTT 126219
QY 4147 AACCAAGTTTAAGCTAGTGTAAATAATTAATAA---AAATGCTTGTCTGTACTTCA 4203
DB 126220 AATTAAGTTTAAGCTAGTGTAAATAATTAATAA---AAATGCTTGTCTGTACTTCA 126279
QY 4204 GCTTGTGTTTATGCCCATTTCAATTTGTTGTCTGTGTGTAATTCATTAACCTTTGATAC 4263
DB 126280 ACTTGTGTTTATGCCCATTTCAATTTGTTGTCTGTGTGTAATTCATTAACCTTTGATAC 126339
QY 4264 ATTTGATGATGTAAGAAATGTTGTTGTAATTAATTAATTAAGAGTTCATTAATTAAT 4323
DB 126340 ATTTGATGATGTAAGAAATGTTGTTGTAATTAATTAATTAAGAGTTCATTAATTAAT 126398
QY 4324 AAATATTTGTGCTGTTAAAAA-----AAATGCTTGTCTGTACTTCA 4358
DB 126399 AAATATTTGTGCTGTTAAAAA-----AAATGCTTGTCTGTACTTCA 126433

RESULT 15
AL513172 142908 bp DNA linear HMG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-291H23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 17 unordered pieces.
ACCESSION AL513172


```

QY 3914 AGGACCTTATTTATTTATGTCCTCTTCTAGGTTAATTCCTCTGATTTGACTTGT 3973
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 429 AGGACCTAATATTTATGTTATGTCCTCTCTAG-----CTGTCTTNGTGTGGAGATTC 375
QY 3974 GAGAGAGAGTTGGACAGTAGATTAGCAAAAGTTCCAAAGTCCAAATTTACAGTGTGAGA 4033
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 GAGGTGTGTTAGAGAGAGAGTTGGAGCAAAAGTTCCAGAGCCGAGAGCAGCTGTGT---G 318
QY 4034 GTGTGGGGGGAATTTAGTCTTATTTTCCCTACATGCGATACACACCTGTG---AATTC 4090
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 GAGTGTGTGGGGAATTTATTTTCCCTACATGCGATACACACCTGTGTGAGATTC 258
QY 4091 AATCTTCAACTGA-----AGGCCCTGCGAGTCTCTCAAAACATAGTT-GTTTGTCTTCT 4144
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 AGCTTCAACAGCAAAAGAGAGCCCTGAGTCTCTCAAAACATAGTTCCCTACTCTTCT 198
QY 4145 TTAAACAAGTTTAAAGTAGTGTATTAATAATTA--AAAAATGCTTGTCTGCTACTTC 4202
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 TTAAACAAGTTTAAAGTAGTGTATTAATAATTA--AAAAAGAAAGCTGTCTGCTCACCTC 138
QY 4203 AGCTTGTGTATGATCCCATTTATTTGTTGCTGTGTGTTAATTTCACTTGTGATAC 4262
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137 AGCTTGTGTATGATCCCGTTTCAATGCTGTGTGCTGTCTAATTCAT-ACITTTGATAC 79
QY 4263 CATTTCTGATGTGTAATAATGTTGTTCTGTATTAATCTTAATAAGAGTTCAATGTAAA 4322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 CATTTCTGATGTGTAATAATGTTGTTCTGTATTAATCTTAATAAGAGTTCAATGTAAA 19
QY 4323 TAAACTATTTGTGGCTGT 4340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18 TAAACTATTTGTGGCTGT 1

```

RESULT 12
AC107643 207214 bp DNA linear HTG 21-APR-2002
LOCUS AC107643
DEFINITION Mus musculus clone RP23-54L14, WORKING DRAFT SEQUENCE, 21 ordered
pieces.
AC107643 2 GI:20258435
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 207214)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-54L14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207214)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B.,
Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McKewan,P., McKernan,K., Meldrim,J., Menus,L., Mhova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

JOURNAL REFERENCE AUTHORS

TITLE JOURNAL COMMENT

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 207214)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
Chazaro,B., Choepeil,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McKewan,P., McKernan,K., Meldrim,J., Menus,L.,
Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2002 this sequence version replaced gl:18308342.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RX/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14249
Center clone name: 54.L.14
----- Summary Statistics
Sequencing vector: plasmid, n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 200414 bases at least Q40
Consensus quality: 203403 bases at least Q30
Consensus quality: 204607 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 205214; sum-of-contigs
Quality coverage: 7.0 in Q20 bases; sum-of-contigs
Quality coverage: 7.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 9177: contig of 9177 bp in length
* 9178 9277: gap of 100 bp
* 9278 9931: contig of 654 bp in length
* 9932 10031: gap of 100 bp
* 10032 10809: contig of 778 bp in length
* 10810 10909: gap of 100 bp
* 10910 13046: contig of 2137 bp in length
* 13047 13146: gap of 100 bp
* 13147 14912: contig of 1766 bp in length
* 14913 15012: gap of 100 bp
* 15013 16787: contig of 1775 bp in length
* 16788 16887: gap of 100 bp

[illegible]

QY 1225 GCCAAGAAACCCGAGAAAGCCCGGTCTTCAGTGGACACCCCAAGACTCTCCACTC 1284
 |||||||
 Db 511 GCCAAGAAACCCGAGAAAGCCCGGTCTTCAGTGGACACCCCAAGACTCTCCACTC 570
 |||||||
 QY 1285 CCCAAAGGGAGCCAGCAGCCCATCCCAAGAGAGAAAGCTTCTCAAGTAGATAGAGATT 1344
 |||||||
 Db 571 CCCAAAGGGAGCCAGCAGCCCATCCCAAGAGAGAAAGCTTCTCAAGTAGATAGAGATT 630
 |||||||
 QY 1345 TGGCGTACCTTCCACACAGAAATGTATTGTGTGCTGGACAGCGCTCCCATCAG 1404
 |||||||
 Db 631 TGGCGTACCTTCCACACAGAAATGTATTGTGTGCTGGACAGCGCTCCCATCAG 690
 |||||||
 QY 1405 CGTTACCATTCAGGGAATCTCTCCAAAGAGAGAGAGACTGTATACAG 1453
 |||||||
 Db 691 CGTTACCATTCAGGGAATCTCTCCAAAGAGAGAGAGACTGTATACAG 739
 |||||||
 RESULT 10
 AC119462/c 193189 bp DNA linear HTG 18-JUL-2002
 LOCUS Rattus norvegicus clone CH230-28N16, *** SEQUENCING IN PROGRESS
 DEFINITION
 AC119462
 AC119462 3 GI:21746590
 VERSION HTG: HTGS_PHASE1.
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 193189)
 Muzny,D.M., Adams,C., Adio-Odola,B., Ali-olsman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbada,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,D., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,K.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsegod,H.,
 Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Mettler,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogun,M., Okunodu,G.,
 Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
 Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Welnsstock,G. and Gibbs,R.
 Direct Submission

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Unpublished
 2 (bases 1 to 193189)
 Worley,K.C.
 Direct Submission
 Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193189)
 Worley,K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20387269.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GURK
 Center clone name: CH230-28N16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 126440 bases at least Q40
 Consensus quality: 140831 bases at least Q30
 Consensus quality: 148852 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 79 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1554: contig of 1554 bp in length
 1555 1654: gap of unknown length
 1655 2715: contig of 1061 bp in length
 2716 2815: gap of unknown length
 2816 3899: contig of 1084 bp in length
 3900 3999: gap of unknown length
 4000 5367: contig of 1368 bp in length
 5368 5467: gap of unknown length
 5468 6690: contig of 1223 bp in length
 6691 6790: gap of unknown length
 6791 8013: contig of 1222 bp in length
 8013 8112: gap of unknown length
 8112 9566: contig of 1454 bp in length
 9567 9666: gap of unknown length
 9667 10914: contig of 1248 bp in length
 10915 11014: gap of unknown length
 11015 12395: contig of 1381 bp in length
 12396 12495: gap of unknown length
 12496 14050: contig of 1555 bp in length
 14051 14150: gap of unknown length
 14151 15648: contig of 1498 bp in length
 15649 15748: gap of unknown length
 15749 16994: contig of 1246 bp in length
 16995 17094: gap of unknown length
 17095 18274: contig of 1180 bp in length
 18275 18374: gap of unknown length
 18375 20426: contig of 2052 bp in length
 20427 20526: gap of unknown length
 20527 22026: contig of 1500 bp in length
 22027 22126: gap of unknown length
 22127 23257: contig of 1131 bp in length
 23258 23358: gap of unknown length
 23359 24652: contig of 1295 bp in length


```

Db      96304 ATCTGCTTAGAGAGATGCTTTCTCTGTGACCGAGAGCTCCCTAGACTTCACCGAGCCAG 96363
Oy      3502 GATGTAGAGAGAAATAGGACTTAATTCACCTAGGGGCTTCATCTCACACCTTAAAGGA -- 3559
Db      96364 GATGTAGAGAGAAATAGGACTTCCTCCACAGTGGCTTATCTCACACCTTAAAGGATG 96423
Oy      3560 -GGAGATTTCTAGAAAACTGGCGCAGATTTCTTGTCTCCATCATTTTAAATGAGCA 3618
Db      96424 GGGATATATCCAGAAATGTGGCGCAGATTTTTTTT -----TTTAAATGTGGCC 96471
Oy      3619 GCGCTTCAGTTTCTTACTCTTACATATGATATTTCTTCGTAAGCTG--TCCAAAA 3676
Db      96472 CACGTATTAGTCCCTACCTGACCTGATTTGTCTTTAATACCTGCTCCATCAAAA 96531
Oy      3677 GAAAAAGAACCAATTCAGTCTCTGCTGCTTGTCTTGTATCC--CTCAGTTCTCTT 3734
Db      96532 ATATATACCTCAATCCGCTGCTCTTACTTTGTTTCTTGATCCCTCAGTTTCTTCTT 96591
Oy      3735 GATTTACAGATGTGTGGGTTCTTAATTTGGGATAGATGAGCAAAATTTAACCATTTGTG 3794
Db      96592 GACAACGAGATTTGATGATGATCTTAA--TTTGGTATGGGTTAATAGACTTAAACGCCAGG 96650
Oy      3795 TTTTGGCCCTACCCAGGAGATCCCGCAGTTTCTGACTTGAAGTACTGAGAAAGATCCA 3854
Db      96651 TTTTGGCCCTAATAGAGAGATCCCTCATATCTG--TTTTGCTAGACTGAGAGATCCA 96709
Oy      3855 CGAGTGTCTATCTGCGCCAGATTTAATAGATTTCTTGGTCTCCCTCCCTCCGCA 3914
Db      96710 TGAGTGTCTATCTGCGCAGATGAGATGATTTCTTGTGTTCT--CTTTCCCTCGA 96768
Oy      3915 GGACCTCTTATTTATTTGTCCTCCCTCTAGTAAATTTCTTGGATTTGACTTTGTTG 3974
Db      96769 GGACCTAATATTTATTTATTTATTTGCTCTCTAGCTGCTT---TTGGTTTGAAGATGCTGA 96824
Oy      3975 AGAAGGAGTTGAGCAGATTTAGCAAAAGTTCCAAAGTTCGAAATTTACAGTGTGTAGAG 4034
Db      96825 GGTGTGTGTGAGAGAGATTTGGAGCAAAAGTTCCAGAGCAGAGCAGAGATGTGT---GG 96881
Oy      4035 TGTGTGGGGAATTTAGTCTTATTTTCCCTACATAGGATGACA--ACACTGTGCA--ATTCA 4092
Db      96882 AGTGTGTGGGAATTTATTTATTTTCCCTACAGTGGATACACAGTGTGTGAGATTCAG 96941
Oy      4093 TCTTCAACTGA-----AGCCCTCAGTGTCTCTTAAACATATGTTGTTTCTTTA 4147
Db      96942 CTTTCCACAGAAAGAGAGGCTCTGCTCTTAAACATATGTTCTTCTTTA 97001
Oy      4148 ACAAGTTAGCTAGTATTAATAATTAATAAAATTTGCTGTCTGCTACTTCAGCTT 4207
Db      97002 --CAAGGTTAAGCTATTAATAATTAATAAAATTTGCTGTCTCC--ACTCAGCTT 97057
Oy      4208 TGTTTTATGCCATTTATTTATTTGCTGTCTGTTGTAATTCATTAATTTGATTAACATTT 4267
Db      97058 TGTTTTATGCCCTGTTTCAATGCTGTCTGTCTGCTGATATTCAT--ACTTTTATACATTT 97116
Oy      4268 CTGATGTGTAATTTGTTGTTCTGTGAATAATCTTATTAAGAGTTCAATTTGTAATAAAC 4327
Db      97117 CTGATGTGTAATTTGTTGTTCTGTGAATAATCTTATTAAGAGTTCAATTTGTAATAAAC 97176
Oy      4328 TATGTGTGCTGTTAA 4342
Db      97177 TATGTGTGCTGTTAA 97191

```

```

RESULT 7
AC131356/c 174346 bp DNA linear HTG 21-AUG-2002
LOCUS      Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS
DEFINITION
***, 55 unordered pieces.
ACCESSION  AC131356
KEYWORDS   HTG; HTGS; PHASE1.
SOURCE     Norway rat.

```

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus

REFERENCE

1 (bases 1 to 174346)
Muzny, D., Marie, M., Metker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benalmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Dubin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Guraratne, P., Healand, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensunewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,
Maneswarth, M., Mahindaratne, M., Mahmud, M., Malloy, K., Mangum, A.,
Manungu, B., Mapua, P., Martin, K., Martin, R., Matliez, E.,
Mawhiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
Newton, N., Nguyen, N., Norris, S., Nwoketeh, O., Okwou, G.,
Olarinunsaagoun, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Perez, A., Perez, L., Pfannkuch, C., Plodder, F., Polndexter, A.,
Popovic, D., Primus, E., Pu, L. L., Puzo, M., Quiroz, J., Rechlin, E.,
Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
Scott, G., Shatman, S., Shen, H., Shetty, J., Shwartsbeyn, A.,
Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E.,
Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
Tindley, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R.,
Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,
Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

COMMENT

Unpublished
2 (bases 1 to 174346)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- genome center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G00J
Center clone name: CH230-278A19
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 1000 of reads
Assembly program: Phrap; version 0.990329

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1547: contig of 1547 bp in length
1548 1647: gap of unknown length
1648 3064: contig of 1417 bp in length
3065 3164: gap of unknown length
3165 4885: contig of 1721 bp in length
4886 6000: gap of unknown length
6001 6100: gap of unknown length
6101 7171: contig of 1071 bp in length
7172 7272: gap of unknown length
7273 8794: contig of 1523 bp in length
8795 8894: gap of unknown length
8895 10368: contig of 1474 bp in length
10369 10468: gap of unknown length
10469 11906: contig of 1438 bp in length
11907 12006: gap of unknown length
12007 13031: contig of 1025 bp in length
13032 13131: gap of unknown length
13132 14575: contig of 1444 bp in length
14576 14676: gap of unknown length
14677 15799: contig of 1024 bp in length
15700 15799: gap of unknown length
15800 16951: contig of 1152 bp in length
16952 17051: gap of unknown length
17052 18695: contig of 1644 bp in length
18696 18795: gap of unknown length
18796 20470: contig of 1675 bp in length
20471 20570: gap of unknown length
20571 22418: contig of 1848 bp in length
22419 22518: gap of unknown length
22519 24364: contig of 1846 bp in length
24365 24464: gap of unknown length
24465 25586: contig of 1122 bp in length
25587 27859: contig of 2173 bp in length
27860 27959: gap of unknown length
27960 29649: contig of 1690 bp in length
29650 29749: gap of unknown length
29750 31331: contig of 1582 bp in length
31332 31431: gap of unknown length
31432 33606: contig of 2175 bp in length
33607 33706: gap of unknown length
33707 35179: contig of 1473 bp in length
35180 35279: gap of unknown length
35280 37193: contig of 1914 bp in length
37194 37293: gap of unknown length
37294 38382: contig of 1089 bp in length
38383 38482: gap of unknown length
38483 40758: contig of 2277 bp in length
40759 40859: gap of unknown length
40860 42763: contig of 1904 bp in length
42764 42863: gap of unknown length
42864 45999: contig of 3136 bp in length
46000 46099: gap of unknown length
46100 49138: contig of 3040 bp in length
49139 49239: gap of unknown length
49240 51762: contig of 2523 bp in length
51763 51862: gap of unknown length
51863 54907: contig of 3045 bp in length
54908 55007: gap of unknown length
55009 57214: contig of 2207 bp in length
57215 57314: gap of unknown length
57315 60299: contig of 2985 bp in length
60300 60399: gap of unknown length
60400 62940: contig of 2541 bp in length
62941 63040: gap of unknown length
63041 65074: contig of 2034 bp in length

65075 65174: gap of unknown length
65175 68290: contig of 3116 bp in length
68291 68390: gap of unknown length
68391 70496: contig of 2106 bp in length
70497 70596: gap of unknown length
70597 73863: contig of 3267 bp in length
73864 73963: gap of unknown length
73964 76872: contig of 2909 bp in length
76873 76972: gap of unknown length
76973 79215: contig of 2243 bp in length
79216 79316: gap of unknown length
79317 83513: contig of 4196 bp in length
83514 83613: gap of unknown length
83614 87309: contig of 3696 bp in length
87310 87409: gap of unknown length
87410 92353: contig of 4944 bp in length
92354 92453: gap of unknown length
92454 95901: contig of 3448 bp in length
95902 96001: gap of unknown length
96002 100934: contig of 4933 bp in length
100935 101035: gap of unknown length
101036 104875: contig of 3841 bp in length
104876 104975: gap of unknown length
104976 109470: contig of 4495 bp in length
109471 109570: gap of unknown length
109571 117676: contig of 8106 bp in length
117677 117776: gap of unknown length
117777 124312: contig of 6536 bp in length
124313 124412: gap of unknown length
124413 128938: contig of 4526 bp in length
128939 129038: gap of unknown length
129039 135586: contig of 6548 bp in length
135587 135686: gap of unknown length
135687 143121: contig of 7435 bp in length
143122 143221: gap of unknown length
143222 149127: contig of 5906 bp in length
149128 149227: gap of unknown length
149228 156513: contig of 7286 bp in length
156514 156613: gap of unknown length

Query Match 11.9% Score 520.6; DB 2; Length 174346;
Best Local Similarity 70.4%; Pred. No. 2.3e-88;
Matches 940; Conservative 0; Mismatches 339; Indels 56; Gaps 22;

QY 3022 CAAGACCTGCTGCTCAATATATAGGAATGCTTCTTACCTTCAGACATATG 3081
||
DB 95899 CACNN 95958
DB 3082 TGAACACAGTAGGGGCTATATCTCCTAGAGGTAGGGCTTTATCTTAAGAGAA 3141
QY 95959 NNN 3142
DB 3142 TGTCCCGAGTATATAGCACTTTAGAGGAGCAAGATATGAGGTGTGCTGG 3201
DB 96019 TGTCCCGGGA-TATATACAC-TTATAGAGGAGCAAGATATGAGGTGTGCTGG 96076
QY 3202 CCCATGAGTGTGACAGAGAGAGAGATGGATACCATTTGCGAGAGAGAAAGTTCC 3261
DB 96077 CCATGTGCTGAGAC---ATTCGAGAAATGGCATTTGCCACGTGGAAGAGCAAGTTTC 96133
QY 3262 TCAGGGGCTCCACACGTCTAAAGTTTGTGAGATGTGATGTGCTTCCGTGATTG 3321
DB 96134 TCA-GGGCCCTCCACACGTCTAAAGCTTTT-TGAGATGT-ATCTGTGCTCCCTGAGATTG 96190
QY 3322 ACTTTAAGAGATTTATCTGAGAGACATGATGATATCTGTGATGATGCTGTGCTT 3381
DB 96191 ACTTTGAAGAGATTTATCCAGCATGACATTTGATGATGATGATGCTGTGCTT 96250
QY 3382 ATTTCTCTTTTGT 3441
DB 96251 AATTCTCTT-----TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 96303
QY 3442 ATCTGCTTAGAGAGTGGCTCTCTATTAAGGAACTGCTGTAACTTCAATTGCAGCAG 3501

```

source
1. 615
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/notes="EMBL/GenBank Accession No. AF102093"
BASE COUNT      84 a      235 c      224 g      71 t      1 others
ORIGIN

Query Match      12.6%  Score 551.6; DB 6; Length 615;
Best Local Similarity 94.6%  Pred. No. 1.3e-94;
Matches 582; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

OY 112 CGGCCCCCTGCGCGGCAATCTGTAGACAGCACTGACTAGAGCGGGCTGCGGCTG 171
    1 CGGCCCCCTGCGCGGCAATCTGTAGACAGCACTGACTAGAGCGGGCTGCGGCTG 60
OY 172 GCGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 231
    61 GCGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
OY 232 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
    121 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
OY 292 CCGGCT---GCGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 348
    181 CTGGCTGCGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
OY 349 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 408
    241 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
OY 409 AAGCGGAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
    301 AAGCGGAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
OY 469 ATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528
    361 ATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
OY 529 CCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
    421 CCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
OY 589 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
    481 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
OY 649 CCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708
    541 CCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
OY 709 GCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
    601 GCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
OY DB

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 COMMENT

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, X., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Geisler, J., Geisler, E., Geisler, K., Gill, R., Grady, M., Guerra, W.,
 Guarnate, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladik, S. L., Hodgson, A., Hughes, M.,
 Hollins, B., Howell, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensbawa, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Maimoud, M., Mallory, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Mundasa, M., Murphy, M., Naik, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwokedeme, O., Okonou, G.,
 Olarinmusa, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polidexter, A.,
 Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S., Sanders, M., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E.,
 Song, X., Z., Sorelle, R., Sosa, J., Stelmach, M., Strong, R., Sutton, A.,
 Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
 Walidron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R.,
 Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 174346)
 Rat Genome Sequencing Consortium.
 Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 -----Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 -----Project Information
 Center project name: G00J
 Center clone name: CH230-278A19
 -----Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 1000 of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 109854 bases at least Q40
 Consensus quality: 118111 bases at least Q30
 Consensus quality: 125098 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 55 contigs. The true order of the pieces

Query Match	Best Local Similarity	Score	DB 2:	Length
Matches	799: Conservative	0: Mismatches	46: Indels	3: Gaps
15	CCGACACCTGGCCCTTCCCAACCCCTCTCCGCTGCGTGCACGGCGCTGCTCCGAGCC	74		
Db	160907	CCGACACCTGGCCCTTCCCAACCCCTCTCCGCTGCGTGCACGGCGCTGCTCCGAGCC	160966	
07	75	ACTATGACCAATGAGATTCGGCGGTGTTCACAGCGCGCGCCCTTCGCGCGGCAATCCT	134	
Db	160967	ACTATGACCAATGAGATTCGGCGGTGTTCACAGCGCGCGCCCTTCGCGCGGCAATCCT	161026	

[illegible]

OY 4340 TAA 4342
 Db 50725 TAA 50723
 RESULT 4
 AC119462
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-28N16, *** SEQUENCING IN PROGRESS
 *** 79 unordered pieces.
 AC119462 193189 bp DNA linear HTG 18-JUL-2002
 AC119462.3 GI:21746590
 HTG: HTGS_PHASE1.
 VERSION Norway rat.
 KEYWORDS Rattus norvegicus
 SOURCE Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Rattus.
 REFERENCE 1 (bases 1 to 193189)
 AUTHORS Murthy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amarutunge,H.C., Are,J.R., Ayala,M., Banks,T.,
 Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonito,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Krtolovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loutsedg,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Mathew,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkwo,S., Ogun,M., Okunolu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
 Soderren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umanal,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 193189)
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193189)
 REFERENCE Direct Submission
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 14, 2002 this sequence version replaced gi:20387269.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GURK
 Center clone name: CH230-28N16

 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye 1008 of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 126440 bases at least Q40
 Consensus quality: 140831 bases at least Q30
 Consensus quality: 148852 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 79 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1554: contig of 1554 bp in length
 * 1555 1654: gap of unknown length
 * 1655 2715: contig of 1061 bp in length
 * 2716 2815: gap of unknown length
 * 2816 3899: contig of 1084 bp in length
 * 3900 3999: gap of unknown length
 * 4000 5367: contig of 1368 bp in length
 * 5368 5468: gap of unknown length
 * 5469 6590: contig of 1223 bp in length
 * 6591 6790: gap of unknown length
 * 6791 8012: contig of 1222 bp in length
 * 8013 8112: gap of unknown length
 * 8113 9566: contig of 1454 bp in length
 * 9567 10914: contig of 1248 bp in length
 * 10915 11014: gap of unknown length
 * 11015 12395: contig of 1381 bp in length
 * 12396 12495: gap of unknown length
 * 12496 14050: contig of 1555 bp in length
 * 14051 14150: gap of unknown length
 * 14151 15648: contig of 1498 bp in length
 * 15649 15748: gap of unknown length
 * 15749 16994: contig of 1246 bp in length
 * 16995 17094: gap of unknown length
 * 17095 18274: contig of 1180 bp in length
 * 18275 18374: gap of unknown length
 * 18375 20426: contig of 2052 bp in length
 * 20427 20526: gap of unknown length
 * 20527 22126: contig of 1500 bp in length
 * 22127 23257: contig of 1131 bp in length
 * 23258 23357: gap of unknown length
 * 23358 24652: contig of 1295 bp in length
 * 24653 24752: gap of unknown length
 * 24753 26187: contig of 1435 bp in length
 * 26188 26287: gap of unknown length
 * 26288 27567: contig of 1280 bp in length
 * 27568 27668: gap of unknown length
 * 27669 28880: contig of 1113 bp in length
 * 28881 29080: gap of unknown length
 * 29081 30136: contig of 1056 bp in length
 * 30137 30236: gap of unknown length
 * 30237 31632: contig of 1396 bp in length
 * 31633 31733: gap of unknown length
 * 31734 33570: contig of 1538 bp in length
 * 33571 33570: gap of unknown length

OY		3464	TCCTPAAAGGGGAACCGCGTAAACCTCAATTCGACAGCAAGGATGTGAGAGAAATAGGACTT	3523
Db		1441	TCCTPAAAGGGAACCTGCTGTAAACCTTCTATGACAGAAAGATGTAGAGAAATAGGACTT	1500
OY		3524	AATTCACATAGGGGCTCTCATCTCACACCTTTAAGAGAGAGATTTCTAGAAAAACTGGGCC	3583
Db		1501	AATTCACATAGGGGCTCTCATCTCACACCTTTAAGAGAGAGATTTCTAGAAAAACTGGGCC	1560
OY		3584	AGATTTTCTTGGTTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTATTCTTACTCTTAC	3643
Db		1561	AGATTTTCTTGGTTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTATTCTTACTCTTAC	1620
OY		3644	CTATGCTGATATTTCTTCGTAACGTCGTCACAAAACAAAAAACCACCAATCACTGCTCTTG	3703
Db		1621	CTATGCTGATATTTCTTCGTAACGTCGTCACAAAACAAAAAACCACCAATCACTGCTCTTG	1680
OY		3704	ACTTTGTTCTTTGATCCCTCAGTTTCTTGTGATTTACAGATGTGTGGGTTCTCAATT	3763
Db		1681	ACTTTGTTCTTTGATCCCTCAGTTTCTTGTGATTTACAGATGTGTGGGTTCTCAATT	1740
OY		3764	TGGGTATGAGTACCAAAATTAACCATGTGTTTGTGCCACACAGGGGACTCCCAAGT	3823
Db		1741	TGGGTATGAGTACCAAAATTAACCATGTGTTTGTGCCACACAGGGGACTCCCAAGT	1800
OY		3824	TTCTGACTTGAAGTAGACTGAGAAAGATCCAGAGGTCTATCTGCCCAGATTTAATAG	3883
Db		1801	TTCTGACTTGAAGTAGACTGAGAAAGATCCAGAGGTCTATCTGCCCAGATTTAATAG	1860
OY		3884	ATTCTAATTTCTTGTTGCCCTGCCCTGAGAGACCTCTATTTTATGTGCCCTCTCT	3943
Db		1861	ATTCTAATTTCTTGTTGCCCTGCCCTGAGAGACCTCTATTTTATGTGCCCTCTCT	1920
OY		3944	AGGTTAATTTCTCTTGATTTGACTTTGTTGAGAAGAGGTTGGACAGTAGATTAGCAAA	4003
Db		1921	AGGTTAATTTCTCTTGATTTGACTTTGTTGAGAAGAGGTTGGACAGTAGATTAGCAAA	1980
OY		4004	GTTCCAAATGCAAAATTAACAGTGTGTGATGTGGGGGGAAATTAATGCTATTATTTTCC	4063
Db		1981	GTTCCAAATGCAAAATTAACAGTGTGTGATGTGGGGGGAAATTAATGCTATTATTTTCC	2040
OY		4064	CTACATGGGATACAAACACTGTGAATTCATCTTCAACTGAAGGCCCTGCACATTCCTTAA	4123
Db		2041	CTACATGGGATACAAACACTGTGAATTCATCTTCAACTGAAGGCCCTGCACATTCCTTAA	2100
OY		4124	AACATAGTGTGTTGTTTCTTTCTTAAACAAGTTTAAGCTAGCTTATATATTAATAAAAAA	4183
Db		2101	AACATAGTGTGTTGTTTCTTTCTTAAACAAGTTTAAGCTAGCTTATATATTAATAAAAAA	2160
OY		4184	TTGCTGTGCTGTACTACCTCAGCTTGTGTTTAATGCCATTTATATATGTGTGCTGTGTTGT	4243
Db		2161	TTGCTGTGCTGTACTACCTCAGCTTGTGTTTAATGCCATTTATATATGTGTGCTGTGTTGT	2220
OY		4244	AATTCATTAACCTTTTGATACCAATTTCTGATGTGTAAATTGGTTGCTTGTAAATATCTTA	4303
Db		2221	AATTCATTAACCTTTTGATACCAATTTCTGATGTGTAAATTGGTTGCTTGTAAATATCTTA	2280
OY		4304	TAAAGAGTTCAATGT 4319	
Db		2281	TAAAGAGTTCAATGT 2296	
RESULT 3				
AL590963/c				
LOCUS	AL590963	196724 bp	DNA	linear
DEFINITION	Mouse DNA sequence from clone RP23-395E10 on chromosome 11, complete sequence.			
ACCESSION	AL590963			
VERSION	AL590963.11	GI:20068458		
KEYWORDS	HTG:			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1
Lovell, J.
Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humayun@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:16304739.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-395E10 is
from the RPI-23 Mouse PAC Library
constructed by the group of Pleter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.

FEATURES
Source
Location/Qualifiers

1..196724
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-395E10"
/clone.lib="RPI-23"

BASE COUNT 49842 a 49408 c 48640 g 48834 t

ORIGIN

Query Match 18.9%; Score 825; DB 10; Length 196724;
Best Local Similarity 69.7%; Pred. No. 2,7e+146;
Matches 1841; Conservative 0; Mismatches 540; Indels 262; Gaps 43;

OY	1832	GAAATTTCAGTACCCTCGGTGATGGATGCCATGCAGATTGGAGATCCGAAGA	1891
Db	53235	GAACTTTGAATACCTCGTGTTGGATGAGCAGACCGCTTGAGATCCGAAGA	5317
OY	1892	GCAAAACACTCACCGAGCGTAGAATAATACCTGCTGCGCAACAACCTGCTTCAGAT	1951
Db	53175	GCACACACCTCACCGGACCTGTAGGAAGTACTGCGTGGC -GGTAGCTGTCTTCAGAT	5311
OY	1952	AGTTGTAGCATGCCATTCC -CGAGAGTGGCAGAGACTGTATATGTGACTTTGCTCA	2010
Db	53116	AGTTGTAGCATGCTGTTCGAGAGATAGCAGAGACGTATGTGTGACCTGTGCCCTCA	5305
OY	2011	CATATGTATACACGCGTGATTAATACCTTCATCACTACTCCCTTGA-----CTTGTGTTTC	2064
Db	53056	TGCTGTTATATTCACACGTGATATGCCCTTTCATCTCCCTTTATATATAGTTTGTTTTC	5299
OY	2065	ATTACTCGATTTACAA-----AAACTCTTTCATTTGGCTAATTTGAGTTA	2112
Db	52996	TTTACTCGATTTACAAAACAAACAGAAAGAACGTCTTGGGTTGGCGATGTGGAATTA	52933
OY	2113	TGAGAGGTGATGGGATTTCTTTTCCTTTTGGGAATATGGGCTCTCAGCTAAAGCTA	2172
Db	52936	--GGGGTTTACTGGAGCGGCTTTTCCCATTTTTGGGGAGATAG--CTTGACACTGCTCTTA	5288
OY	2173	TAGAGTGGCATTACAGAGATTTCAGGGCGCTGTTTCTATACATTGGCTATGTTAAAGG	2232
Db	52881	TAGG--GGGAACCTTGGGGGTCACAGGGGCCACTTTGATCACTGCTGCTCCTAAGCGG	5282
OY	2233	GGTAAAGGCGCTCTTCATTATGACATCTGGAAGATGAAGCAGCCCTTCCTTAGAGCT	2292

OY 2252 TTAGACATGTGGAAGATGAGACGCCCTTCCTTTAGAGCTGTGCTGCATGGCCTCT 2311
|||||
Db 78303 TTAGCATGTGGAAGATGAGACGCCCTTCCTTTAGAGCTGTGCTGCATGGCCTCT 78382
OY 2312 CTCACCCGTGTACACCCCTCTTATAGTGGATATAGTATTTTAAACCTTAATAAACA 2371
|||||
Db 78363 CTCACCCGTGTACACCCCTCTTATAGTGGATATAGTATTTTAAACCTTAATAAACA 78422
OY 2372 AACAACTTCACCATAGCTTTAGACCCAGAGAGAAATGACAGATGAAAGCCATGACAA 2431
78423 AACAACTTCACCATAGCTTTAGACCCAGAGAGAAATGACAGATGAAAGCCATGACAA 78482
OY 2432 GCCATCTTCACAGATAGAAAAAGACATGAGAGTGTGATTAATCTGTGAAGATA 2491
78483 GCCATCTTCACAGATAGAAAAAGACATGAGAGTGTGATTAATCTGTGAAGATA 78542
OY 2492 GTTGTTCATTTGAAACTATCTGTGATACAGTATGTGGAAAGGATGTTTGGCTGTAT 2551
78543 GTTGTTCATTTGAAACTATCTGTGATACAGTATGTGGAAAGGATGTTTGGCTGTAT 78602
OY 2552 TATTTTTCAGTATATGATATACATTTCTTACGTCCAAAACCAAAATCTTTGAAA 2611
78603 TATTTTTCAGTATATGATATACATTTCTTACGTCCAAAACCAAAATCTTTGAAA 78662
OY 2612 AGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCTGTAAACAAAAGCACTAGTAT 2671
78663 AGAAAGTGGGATGTTAGTTTCAGAACAGTTACAGCTGTAAACAAAAGCACTAGTAT 78722
OY 2672 TTGGATGTGCATGCCAAAACCTGTATAATGTCTTATACATCACTCTTCAAGTAT 2731
78723 TTGGATGTGCATGCCAAAACCTGTATAATGTCTTATACATCACTCTTCAAGTAT 78782
OY 2732 CCTTCATTTGGGCTTCATCTTTTACAGAACCTGTGGTGGGATGAGCTTGGGG 2791
78783 CCTTCATTTGGGCTTCATCTTTTACAGAACCTGTGGTGGGATGAGCTTGGGG 78842
OY 2792 GGTAGGGGAGAGTGTGAAATAGTGTCTCTTGGCTGGCAATGTCTACATCTTGA 2851
78843 GGTAGGGGAGAGTGTGAAATAGTGTCTCTTGGCTGGCAATGTCTACATCTTGA 78902
OY 2852 ACAAAACATATGACTTAATGAGCTTCTCATTGCTTGTAAATAATTTGTATGTGA 2911
78903 ACAAAACATATGACTTAATGAGCTTCTCATTGCTTGTAAATAATTTGTATGTGA 78962
OY 2912 CCATCTTGTCTCCTCCCTCCCTGTTTGTAAATATCAGAGTATGCACTCCAGGCCAC 2971
78963 CCATCTTGTCTCCTCCCTCCCTGTTTGTAAATATCAGAGTATGCACTCCAGGCCAC 79022
OY 2972 TTTGCTCAGTGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAAAGCTCT 3031
79023 TTTGCTCAGTGTAGATCCCTATTAATCTGAAAGAAATAGAGCCAAAGCTCT 79082
OY 3032 GGTCTCAATATATAGGAATGCTTCTTATAGTCTTACAGACTTGTGTGAAACAAG 3091
79083 GGTCTCAATATATAGGAATGCTTCTTATAGTCTTACAGACTTGTGTGAAACAAG 79142
OY 3092 TAGGGCTTAATCTCTAGAAAGTGGGGCTTTATCTTAAAGAAATATGTCCCGAGA 3151
79143 TAGGGCTTAATCTCTAGAAAGTGGGGCTTTATCTTAAAGAAATATGTCCCGAGA 79202
OY 3152 TTATTTAGCACTTTTATAGAGAGAAAGCAAGTATAGAGGTGTGTGCTGGCCACAGT 3211
79203 TTATTTAGCACTTTTATAGAGAGAAAGCAAGTATAGAGGTGTGTGCTGGCCACAGT 79262
OY 3212 GAGCAGCAAGAGAGATGGATACCATTTGTGGAGAGAGAAAGTTCTCAGGGCCT 3271
79263 GAGCAGCAAGAGAGATGGATACCATTTGTGGAGAGAGAAAGTTCTCAGGGCCT 79322
OY 3272 CCCACTGCTAAAGTTTTTGTGAGATGTGATCTGTGCTCTGATTTGACTTTTAAG 3331
79323 CCCACTGCTAAAGTTTTTGTGAGATGTGATCTGTGCTCTGATTTGACTTTTAAG 79382

OY 3332 GAATATTTCTGGACACACATGATGATTTCTTGATGATCTTGTCTCTTATTTCTCCT 3391
|||||
Db 79383 GAATATTTCTGGACACACATGATGATTTCTTGATGATCTTGTCTCTTATTTCTCCT 79442
OY 3392 TTGT 3451
79443 TTGT 79502
OY 3452 CAGAGTGGGCTCTCTATTAAGGAACCTGTGTAACTTCATTTGACGACGAAGATGTAGA 3511
79503 CAGAGTGGGCTCTCTATTAAGGAACCTGTGTAACTTCATTTGACGACGAAGATGTAGA 79562
OY 3512 GAATATGAGCACTTAATTCACATGAGGGCTCTCATCTCACACCTTAAGAGAGATTTAG 3571
79563 GAATATGAGCACTTAATTCACATGAGGGCTCTCATCTCACACCTTAAGAGAGATTTAG 79622
OY 3572 AAAAAGTGGCCAGATTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3631
79623 AAAAAGTGGCCAGATTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 79682
OY 3632 TCTTACTCTGTACCTATGTATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3691
79683 TCTTACTCTGTACCTATGTATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 79742
OY 3692 CAGTGTCTGTGACTTGT 3751
79743 CAGTGTCTGTGACTTGT 79802
OY 3752 GGTTCCTAATTTTGGGTATGAGTTAGCAAAATTTAACCATTTGTGTGTGTGTGTGT 3811
79803 GGTTCCTAATTTTGGGTATGAGTTAGCAAAATTTAACCATTTGTGTGTGTGTGTGT 79862
OY 3812 GGACTGCCAGTTTGTGACTGTGAAGTATGAGTATGAGTATGAGTATGAGTATGAGT 3871
79863 GGACTGCCAGTTTGTGACTGTGAAGTATGAGTATGAGTATGAGTATGAGTATGAGT 79922
OY 3872 AGATTTAAGTATGATTTCTATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931
79923 AGATTTAAGTATGATTTCTATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 79982
OY 3932 GTCCCTCTCTGAGTTAATTTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3991
79983 GTCCCTCTCTGAGTTAATTTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 80042
OY 3992 TAGATTACCAAGTTCCAAGTGCACAAATTTACAGTGTGTAGAGTGTGGGGGAAATAG 4051
80043 TAGATTACCAAGTTCCAAGTGCACAAATTTACAGTGTGTAGAGTGTGGGGGAAATAG 80102
OY 4052 TCTTATTTTCTCCATGATGGATACAGACTGTGAATTCATTTCAACTGAAAGCCCTG 4111
80103 TCTTATTTTCTCCATGATGGATACAGACTGTGAATTCATTTCAACTGAAAGCCCTG 80162
OY 4112 CAGTCTCTCTAATCATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4171
80163 CAGTCTCTCTAATCATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 80222
OY 4172 AATTAATAAATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4231
80223 AATTAATAAATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 80282
OY 4232 TGTCTGT 4291
80283 TGTCTGT 80342
OY 4292 GTAATATCTTATTAAGAGTTCATTTGTAAATTAATTTGTGTGTGTGTGTGT 4342
80343 GTAATATCTTATTAAGAGTTCATTTGTAAATTAATTTGTGTGTGTGTGTGTGT 80393
Db 80393

RESULT 2
AK055378
LOCUS AK055378 2296 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ30816 f1s, clone FBRA2001571.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:18 ; Search time 10970 Seconds

(without alignments)
11598.675 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372

Sequence: 1 cagcctcgcagcccccgcac.....aaaaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY-NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2511	57.4	180017	2	AC068669	AC068669 Homo sapi
2	2294.4	52.5	2296	2	AK055378	AK055378 Homo sapi
3	825	18.9	196724	10	AL590963	AL590963 Mouse DNA
4	758.4	17.3	193189	2	AC119462	AC119462 Rattus no
5	551.6	12.6	615	6	AX401049	AX401049 Sequence
6	520.6	11.9	174346	2	AC131356	AC131356 Rattus no
7	520	11.9	174346	2	AC131356	AC131356 Rattus no
8	510.2	11.7	542	6	AX368844	AX368844 Sequence
9	380.2	8.7	1433	9	HS8800235	AL049450 Homo sapi
10	371	8.5	193189	2	AC119462	AC119462 Rattus no
11	274.8	6.3	590	6	AX400340	AX400340 Sequence
12	242	5.5	207214	2	AC107643	AC107643 Mus muscu
13	240	5.5	240	11	G22752	G22752 human STS W
14	235.4	5.4	135060	9	AL354659	AL354659 Human DNA
15	235.4	5.4	142808	2	AL513172	AL513172 Homo sapi
16	235.4	5.4	165311	9	AC097061	AC097061 Homo sapi
17	213.2	4.9	177607	2	AC119484	AC119484 Rattus no
18	212	4.8	171318	2	AC127122	AC127122 Rattus no
19	211.2	4.8	201143	2	AC125268	AC125268 Mus muscu
20	166	3.8	207214	2	AC107643	AC107643 Mus muscu
21	150.2	3.4	144577	2	AC123367	AC123367 Rattus no
22	107.2	2.5	565	6	AX401061	AX401061 Sequence
23	92.6	2.1	72397	2	AC114443	AC114443 Rattus no
24	90	2.1	132727	2	AC120736	AC120736 Rattus no
25	87.4	2.0	175440	2	AC126197	AC126197 Rattus no
26	87.2	2.0	74721	2	AC096161	AC096161 Rattus no
27	86.6	2.0	991	11	PM128128	AL684455 Penicilliu
28	85.2	1.9	184402	2	AC127041	AC127041 Rattus no
29	84.2	1.9	885	11	PM7F8G	AL685189 Penicilliu
30	83.8	1.9	55802	2	AC103271	AC103271 Rattus no
31	83.8	1.9	150090	2	AC121750	AC121750 Rattus no
32	83.4	1.9	202509	2	AC129875	AC129875 Rattus no
33	83.2	1.9	1094	11	PM7G11B	AL685196 Penicilliu
34	83.2	1.9	145592	2	AC119341	AC119341 Rattus no
35	82.8	1.9	300695	2	AC079431	AC079431 Mus muscu
36	81.8	1.9	26085	2	AC096530	AC096530 Rattus no
37	81.6	1.9	61958	2	AC121467	AC121467 Rattus no
38	81.6	1.9	101282	2	AC121474	AC121474 Rattus no
39	81.6	1.9	195227	2	AC129795	AC129795 Rattus no
40	81.4	1.9	125020	9	AR429315	AR429315 Homo sapi
41	80.8	1.8	69217	2	AC129516	AC129516 Homo sapi
42	80.8	1.8	136014	2	AC126531	AC126531 Rattus no
43	80.8	1.8	174410	2	AC128158	AC128158 Rattus no
44	80.6	1.8	179608	2	AC128497	AC128497 Rattus no
45	80.4	1.8	1065	11	PM2B12B	AL684695 Penicilliu

ALIGNMENTS

```
RESULT 1
LOCUS      AC068669      180017 bp      DNA      HTG 16-FEB-2002
DEFINITION Homo sapiens chromosome 17 clone RP11-749116 map 17, *** SEQUENCING
IN PROGRESS ***; 16 unordered pieces.
ACCESSION  AC068669
KEYWORDS   HTG; HTGS_PHASE1; HTGS_FUL1TOP; HTGS_ACTIVERIN.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 180017)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 17, clone RP11-749116
```


QY 249 CCCTGCTCTCTCCAGGCGGGAGCCCGCCCTTCCCGCGCGCTCCCGCGGCAAG 308
|||||
Db 841390 CCGC-----CATCACAGTGGGAGCGTCCCGCGCTTCCCGCGCGCGCGCGCG 841337
QY 309 GCGCGGGGCTTACTTCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 368
|||||
Db 841336 CCAGTGAAGTGGCCGCGCGCGCTCCCGCTTCCCGCGCGCGCGCGCGCG 841277
QY 369 GGTTCGGTGCCTTCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 428
|||||
Db 841276 GCG 841217
QY 429 CTTCCGCGAGCGCGCGCTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 488
|||||
Db 841216 CCGACCG 841163
QY 489 CAGACGGGCTCTCTGTTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 548
|||||
Db 841162 CCACCG 841103
QY 549 GGGCGGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608
|||||
Db 841102 CCG 841043
QY 609 CCGCGGCGCG-----CCACCGCTTCCCGCGCGCGCGCGCGCGCGCGCG 659
|||||
Db 841042 CCGTTCGCGATCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840987

RESULT 15

US-08-984-709A-49
; Sequence 49, Application US/08984709A
; Patent No. 6320032
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark E.
; APPLICANT: Stauderman, Kenneth A.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, Suite 700
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,709A
; FILING DATE: 02-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 450-8400
; TELEFAX: (619) 587-5360
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7898 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:

ORIGINAL SOURCE:

; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 249...7307
; OTHER INFORMATION:
; US-08-984-709A-49

Query Match 1.3%; Score 55.6; DB 4; Length 7898;
Best Local Similarity 46.7%; Pred. No. 0.004;

Matches 251; Conservative 0; Mismatches 279; Indels 8; Gaps 2;
QY 102 AAGCG 161
|||||
Db 3 AGCG 62
QY 162 GCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 221
|||||
Db 63 CCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
QY 222 CACCGTAAGCTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 281
|||||
Db 123 GCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
QY 282 CTTTCCCG 341
|||||
Db 183 CG 242
QY 342 CCGCGCGCAGCAGGAGAGAGCTGGG--GCGGTTGCGTGGCGCGCGCG 399
|||||
Db 243 GCCACCATGACCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 302
QY 400 CCACCAAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
|||||
Db 303 CCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
QY 460 GCGCGAAGTATCAGCGCGTGTCTCCCGCATTCAGACGGGCTCTCTG 519
|||||
Db 357 GGACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
QY 520 AGCCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
|||||
Db 417 GAGCG 476
QY 580 ACCCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 637
|||||
Db 477 GCCACGCTCTTCTTCTGCGTCCGTCAGACCGCGCGCGCGCGCGCGCG 534

Search completed: July 14, 2003, 13:41:47
Job time : 197 secs

OY	331	CCGGGGGGGGCCCCGGGCGAGCGAAGAGACTGTGGGGCGGTTCGTGCTCCCTTGGCCCTGTGCT	390
Db	4208	CCAGCTCCGGACACCAAGTTCGGGCTCTCCGGCTCTCTCCGGCTACGCCAGCTCCCGCCGGC	4149
OY	391	CGCCCCGGGGCCACCAAGCAAGCCGGCATTTGGGGGGGAGCCCTGCCGCA6CCGGAGCCGGCT	450
Db	4148	CCGGCCCGCGAGCCACGAGCCGCCAGCGGGCGCAAGAAAGAGCCGCGCGCGCCGGGGCCGGCC	4089
OY	451	GCAGCCCCCGGCCCAAGATATGAGCGGGTCT-GGCCATTCTGAGCGGGGTCTCTGTGGCG	509
Db	4088	GCCCGGGGGAGGGGAGGAGGACGAGGGGCTCTCCGGCTCGGCGCTCCGGGGGACGGCC	4029
OY	510	GCGGCCAAAGAGCTTACGCCCTTGGGCTGTGGGACAAAGGTGGGGCGGCTCCCGCGTGGC	569
Db	4028	ACGGCCACCGGGAAGACAAAGAGACCGGGGGCCCCCGCGCAAGCGCGCTCCCTCGGAC	3969
OY	570	ACCCGCTCGAGCCCGCGGGGAGCCCGCAACATACCTTGCACGGGGGCGCAACCCCTCGCG	629
Db	3968	TGGGGCCCGGCTCGGAGCCCGCTCCCGCGCTCTGATGTCTCTCTCTCTCTCTCTCTCT	3909
OY	630	CCCAACCGCACCGCC	644
Db	3908	CCGAGGAGCAGCGGC	3894

RESULT 13
 US-07-867-106-2/c
 Sequence 2, Application US/07867106
 Patent No. 5389526
 GENERAL INFORMATION:
 APPLICANT: Slade, Martin B
 APPLICANT: Chang, Andy C M
 APPLICANT: Williams, Keith L
 TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
 TITLE OF INVENTION: Slime Moulds of the genus Dictyostellium
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526RIS
 STREET: One Liberty Place 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/867,106
 FILING DATE: 19920625
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PJ 7187
 APPLICATION NUMBER: PCT/AU90/00530
 FILING DATE: 02-NOV-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Feeney, Joanne Longo
 REGISTRATION NUMBER: 35,134
 REFERENCE/DOCKET NUMBER: RICE-0002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5852 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: CDS

```

; LOCATION: 2378..5038
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
;
US-07-867-106-2

```

```
Query Match      1.3%; Score 56.4; DB 1; Length 5852;
Best Local Similarity 53.1%; Pred. No. 0.0023;
Matches 120; Conservative 0; Mismatches 106; Indels 0; Gaps 0
```

[illegible]

RESULT 14

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A

; GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCING

```

: TITLE OF INVENTION: DNA SEQUENCES FOR SRAIN ANALYSIS IN MICROBACILLARIUM
:
: TITLE OF INVENTION: TUBERCULOSIS
:
: FILE REFERENCE: 24366-20007 00

```

FILE REFERENCE: 24300 20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24

```

/ 0 SOFTWARE: PatentIn Ver. 3.1
; NUMBER OF SEQ ID NOS: 2

```

```

; SEQ ID NO 2
; LENGTH: 4403765

```

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

```

```

;
; FEATURE:
; OTHER INFORMATION: CDC 1551
;

```

```

; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g

```

US-09-103-840A-2

Query Match	Score	DB	Length
1.3%	56	4	4403765

Best Local Similarity 46.3%; Pred. NO. 0.069;
Matches 304; Conservative 0; Mismatches 335; Indels 17; Gaps 3.

9 GACCCCCCGACCTCGCGCCCTTCCGCCACCCCTCTCTCGGCTCGGTGCCCGGCGCTGCTC 68

Db
841630 G C C C C G C C G G G C C A C C G T C C C C G C C G C C G G T G C T C G C G C C G C C G T T G C C G C C G T T G C C G 84157

69 CGGACCACTATGACCATGACATCCGCGGTTCAGGCGGCGCGCGCCCTGCCGCGGC 128

Db 841570 CCGTTCGGATCAACCCGGGGGCCCGGCCCCCGGCGGGCGGCCTCCTCCGAGCGTT 841511

QY 129 AATCCTGACGACGACTGCATACGACGGGCGCTGGCGGCGGCCGAGACGAG 188

Db
841510 CTACTAATGCGCCGCCGCCGCCGGAGCCGCCGCCGGAACACAATGCCGGCGTCGCCG 84145

189 CCTGGGCGGGCCGAGCCCACTTCTCTCCCGCCGACCGTAACTCAAGGAGCGGGGCC 248

D8 841450 CCGGCTCCGGCTGCCCCGGCGAATCCGGCGCTGCGACTAGCGCCGGCCGGTCCCGCGGGCC 84139

QY	257	TTCCGCCAGGAGGAGCCCGGCGCTTTCCCGCCGGCGCTGGCGCGCAAGAGCGCGGG	316
Db	928944	CGCCGCCAGACAGCCCGGTCCCGCTGCTCCGGCGCGCGTTGCGCGGTGCGGTG	
QY	317	CTTGTTACTCCGCGCGGCGGCGCCCGCGGCGAGCAGAGAGAGCTGGGCGGTTTCG	376
Db	929004	CGGCTCCGCAACCCCGCCGAGCGCGCGATCGCGGCGCGCATTCGAGAGGATCG	929063
QY	377	GCCTTGGCCCTGTCCGCCCCGGGCCACCAAGACCGGCATTGGGGGGAGCCTCCG	436
Db	929064	CGCGGTCCTCCCATGGCCGCGCTACACGCGAAGCGCGCGCGCGGACCGCCGCG	929123
QY	437	AGCGGAGACCGGCTGACGCCCGCGCCCAATGATAGGCGGTGTCGCCATTGCAGCG	496
Db	929124	CGCGGCGCGCGCC--CAGCAGCGCGACGCCCAATCGCGCGCGCGCCATGCGCGGT	929181
QY	497	CTCTCTGTCGGCGCGCGCAAGACCTTACGCCCTGGGCTGGGGACAAAGTGGGCGG	556
Db	929182	CTCGGTGGCGGCCCGCGCACGCCCGCGCGCGCGCGCGAGCGCCACCGCAAGACCGCG	929241
QY	557	CTCCCGCGCTGCACCGGCTGCGGACCGCGCGGAGCCCGCACCACTTCTCTGCGCGG	616
Db	929242	CGCGCGCGCACCAAGCGGACCGGTCCCGCGGTGCGCGGCGACCGGTGCAAGCCCG	929301
QY	617	GCCACCCCTCGCGCCACCGCACCGCACCGGAGACCTTGGCGGCGACGAGGCGAGATG	676
Db	929302	AAGCGCGCGCGCGCGGTAGCGCGCGCGCAACCGATGAAAGTCGGGGGCAACACG	929361
QY	677	G 677	
Db	929362	G 929362	

```

RESULT 10
US-09-105-537-1
: Sequence 1, Application US/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D. H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.4380U1
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 15872
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	1.3%;	Score 57.6;	DB 4;	Length 15872;
Best Local Similarity	46.1%;	Pred. No. 0.0019;		
Matches 298;	Conservative	0;	Mismatches 344;	Indels 4;
			Gaps	3

QY	13	CCCCGACACCTGGCCCTCTCCCAACCCCTCTCGGCTGGGTGCCGGGGCTGTCTCCGA	72
Db	11368	CCCCGTACCCCTCCCGTGAGACCTCCGGCCCGCGACCGGGGCTCTTTCGGGGCACTGCT	11427
QY	73	CCACTATGACCATGAGATCCGGGGTGTTCAAAGCGGGCCGGGCCCTTGGCCGGGGCAATC	132
Db	11428	GCGGTCCG-CCACCGGACCTGCACACAGAGGTGACACACCGCGGCCGTGGCCGGGTGC	11486
QY	133	CTAGACGACCACTGAGCTATGACAGCGGGCTCGGGCTGTGGCGGGCCCGAGAGAGACTTG	192
Db	11487	TGTGCTGTCTGTCCGAGAGAGATGGGGCCCATGCTCAGACAGCGCCCGGTACCCGCGGGG	11546
QY	193	GGGGGGCCGAAGCCCACTTCTCCCCCGGACCGCTAAGCTCAAGAGACCGGGGGCCCCCGC	252

Db	11547	TCCTGGGAGAGCTGTCCCTGATGCAAGCTATGAGAGGAGGAGCGGTGAGAGGCTCGCGCTGT	11606
OY	253	TGCGCTCTCTCCCAAGGCGGGAGGCCCGCGGCTTCCCGCGCGGCTCGCGCGCAAGGCC	312
Db	11607	GGTCCGCTCTCCCGCGCGCGCGGTGCGCGCCCGACCGGGAACGCCCGTCCGCGCGGGG	11666
OY	313	GGGCGCTGTACTCCCGCGCGGGGCGGCGCCCGGAGCAGGAAAGACGTGTGCGCGGTT	372
Db	11667	CCGCGCTGTGGGGGGCTGAGGGGGGTGAGCGCCCTTGAGACGCCCAACCGGTGGGCGGTTC	11726
OY	373	CGGTGCCCTTGCCTGTCCGCGCCCGCGACCAAGCAAGCGCGCATTTGGGGGGAGACCTG	432
Db	11727	TCGTGAGACTGCCCGCGCTCCCGGATGCGAGCGACTGGGGGCGCGCTGGAAACGCGTGG	11786
OY	433	CCGCAAGCC--GAGAGCGGCTGACACCCCGGCGCCAGTATCAGGGCGTCTGCCATTACA	490
Db	11787	CCGGTCTCCGAGGACCAATGCGCGTGGCGCGTCCGCGAGTGGGGCGCGGCGCTTCACCA	11846
OY	491	GACGGGCTCTACTCGTGGCGGCGGCGCAAAAGAGCTTACGCCCTGGGGCTGGGAGACAAGGGTG	550
Db	11847	GGCTGCGCGCGC-GACGGCGGCGGGCGCGAGCGCGCAACCGCGTACCGGCGCGCGGCAAG	11905
OY	551	GGCGGCGCTCCCGCGTGCACGCGCTTGAGACCGGCGGGAGCCGCCACCACTTACTTCGCC	610
Db	11906	GTCGTCGTCAACCGGTGACACGGCGCGCTCGGCGGGCAATCTCCCGCTGCGCTCGCGCG	11966
OY	611	CGGGCGGCGCAACCTCGCGCGCCACCGCAACCGCGCGGAGACCTTGGCG	676
Db	11966	GCGGGCGCGCAACCTGCGCGCTACCGACCGCGCGGGCGCCCGGACG	12011

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```



```

; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F15
;
US-08-232-463-14

Query Match 1.5%; Score 67; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. No. 8.6e-06;
Matches 10; Conservative 233; Mismatches 138; Indels 0; Gaps 0;

QY 784 AGCAGACAGAGCTGCGCAAGAAAGAGAGATCGAGAGCTGAAGTCAGAGAGACA 843
DB 1432 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1373
QY 844 CGCTCTGCTGCTGAGTGAAGTGAAGGAGGAGTGCAGTGTAAAGAGATTAACG 903
DB 1372 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1313
QY 904 AGAAGAAAGGACAGAGCTGTTGACGGCTATGAAGTGAAGAGAGAGAAAGACAGC 963
DB 1312 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1253
QY 964 TATCTGAGAAATTAAGTGAAGTGCAGCGGAGCTTCCGAGACATCCAGACTCTGC 1023
DB 1252 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1193
QY 1024 CTCCTCAAGCCCTTCTATGTGGCGGAGTGGAAAGAGACATAAAGAAATCCCATTTG 1083
DB 1192 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1133
QY 1084 GAAGTCAGAAAGAGAGCTCCTGTAAGAAAGCTGCTCCTGTAATTTCAAAAGTCAAAA 1143
DB 1132 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1073
QY 1144 CAAAACTCTAGACACTCTC 1164
DB 1072 RRRRRATCGCAAGCTCCCTC 1052

RESULT 6
US-09-298-568-3/c
; Sequence 3, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
```

```

; APPLICANT: Bailestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
;
US-09-298-568-3

Query Match 1.4%; Score 62; DB 4; Length 801;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;
Matches 242; Conservative 0; Mismatches 235; Indels 9; Gaps 3;

QY 236 GAGCGCGGGCCCCCGCTGCTCCAGAGGCGGGAGCGCCCGCTTCCCGGCGCG 295
DB 512 GGGACAGGGGGCCCCCGCCGAGCGCCCGGGAGGGGGCGCCCGCGCGCGG 453
QY 296 CTGCGCGCGCAAGGCGCGGGGCTTGTACTCCCGCGGGGCGCCCGCGGAGAGA 355
DB 452 CTGCGCGCGCGGGGGCGCGGAGAAAGCAAGCGCCCGCGCGCGGGCGG-- 395
QY 356 AGAGAGCTGGGGGGGTTGGTCCCTTGCCCGCCCGCGCGGCAAGCAAGCAAGCGG 415
DB 394 ---CGCGCGGGCGGATCCCGCCCGCTGTTTACCCCGCGGGGGCGCGCGG 338
QY 416 CATTTGGGGGGAGACCTTGGCGGAGCGGAGCGGCTCAGCCCGCGCGCAAGTATAGGC 475
DB 337 CGCGGGGAGACGGGGGGCGAGAGAGGTGCGGGGGCGCGCGCAAAAAAAGTAGGCGT 278
QY 476 GGTGTGCTCCATTCACAGCGGCGTCTCTGTCGGCGGCGCAAGAGCTTACGCTGGGC 535
DB 277 GGGCTAGCCCGTACCCCGCGGGGCTCGGGGCTCGGGGCGCTCGCGCGCGC 218
QY 536 TGGGGACAAAGGTGGGGCGGCTTCCCGCTGCGACCGC--CTCGAGCCGGCGGAGC 592
DB 217 TGGCGGCGCGGGGGCGGCTCGCGCTCGCTCGCGCGGGCGGGGGCGCGCGCGC 158
QY 593 CCCAGCACTACTCTGCGCGGGCGGCGGCGGCGC--CTCGCGCGCGGCGGCGGAGCC 651
DB 157 CTCCCGCGTCCCGCGCGCGGAGAGCGCGCGGCGGCGCTTGCCTCCCGCGGCGG 98
QY 652 TGGCGGCGAGGCGAGATGGAAGATGGAAGAGCCCTCTCGGGGGTGTGGCG 711
DB 97 CGGCTCGGGCGGGGCGTGGCGGAGCGCGCGGAGCGAGCGGGGGAGGGAGGGGCG 38
QY 712 GCTCGG 717
DB 37 GCGCGG 32

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

LENGTH: 801
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-16

Query Match 1.6%; Score 70; DB 4; Length 801;
Best Local Similarity 50.8%; Pred. No. 5.9e-07;
Matches 247; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

```
QY 236 GGAGCGCGCGCGCGCGCGCTGCGCTCTCCAGGCGCGGAGCCCGCGCTTCCCGCGCGG 295
    || || || || || || || || || || || || || || || || || || || || ||
DB 512 GGGACAGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 453
QY 296 CTGCGGCGCGCAAGGCGCGCGCGCTTGTACTCCGCGCGCGCGCGCGCGCGCGCGCGG 355
    || || || || || || || || || || || || || || || || || || || || ||
DB 452 CTCGCGCGCGCGCGCGCGCGCGCGGAGAAAGCAAGCAAGCGCGCGCGCGCGCGCGG 395
QY 356 AGAGAGCTGGGCGCGCGCTGCGTCCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGG 415
    || || || || || || || || || || || || || || || || || || || || ||
DB 394 ---CGCGCGCGCGCGCATCCCGCGCGCTGTTTACCGCGCGCGCGCGCGCGCGCGG 338
QY 416 CATGGGCGGAGCGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 475
    || || || || || || || || || || || || || || || || || || || || ||
DB 337 CCGCGGCGGACCGCGCGCGCGCGAGAGGTTGCGGCGCGCGCGCGCGCGCGCGCGG 278
QY 476 GGTGCTGCGCATTCAGACGCGCGCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGG 535
    || || || || || || || || || || || || || || || || || || || || ||
DB 277 GCGCTAGCGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 218
QY 536 TGGGACAAAGGTTGGGCGCGCGCTCCCGCGCTGCCACCGC---CTCGGACCGCGCGG 592
    || || || || || || || || || || || || || || || || || || || || ||
DB 217 TGGGCGCGCGCGCGCGCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 158
QY 593 CCCACGACTACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 651
    || || || || || || || || || || || || || || || || || || || || ||
DB 157 CTCCCGCGCTCCCGCGCGCGCGCGCGAGAGACCGCGCGCGCGCGCGCGCGCGCGG 98
QY 652 TGGCGCGCGCGCGCGCGCGAGATGAGAGATATGAGAGAGCGCTCTCGGCGGTTGGCG 711
    || || || || || || || || || || || || || || || || || || || || ||
DB 97 CGGCTCGCGCGCGCGCGCTGCGCGAGCGCGCGGAGCGCGAGCGGCGGAGAGGCGG 38
QY 712 GCTCGG 717
    || || ||
DB 37 GCGCGG 32
```

RESULT 4

US-08-018-977C-4
Sequence 4, Application US/08018977C

Patent No. 5686601

GENERAL INFORMATION:

APPLICANT: Weber, Peter C.

TITLE OF INVENTION: DNA Sequences Encoding Mutant Antiviral

TITLE OF INVENTION: Regulatory Proteins

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: The Pennsylvania State University

STREET: 113 Technology Center

CITY: University Park

STATE: PA

COUNTRY: US

ZIP: 16802-7000

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 720 Kb storage

OPERATING SYSTEM: MS-DOS 6.22

SOFTWARE: Microsoft word for Windows 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/018, 977C

FILING DATE: 18-FEB-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/726, 071

FILING DATE: 05-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Thomas J. Monahan
REGISTRATION NUMBER: 29, 835
REFERENCE/DOCKET NUMBER: 91-1039/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (814) 865-6277
TELEFAX: (814) 865-3591
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 936 bases
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-018-977C-4

Query Match 1.5%; Score 67.2; DB 1; Length 936;
Best Local Similarity 46.5%; Pred. No. 2.9e-06;
Matches 251; Conservative 0; Mismatches 288; Indels 1; Gaps 1;

```
QY 86 GAGATCCCGGCTTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 145
    || || || || || || || || || || || || || || || || || || || || ||
DB 213 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 272
QY 146 GACTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 205
    || || || || || || || || || || || || || || || || || || || || ||
DB 273 GGTGTTGAACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 332
QY 206 CCACTTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 265
    || || || || || || || || || || || || || || || || || || || || ||
DB 333 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 392
QY 266 GGGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 325
    || || || || || || || || || || || || || || || || || || || || ||
DB 393 CCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 452
QY 326 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 385
    || || || || || || || || || || || || || || || || || || || || ||
DB 453 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 512
QY 386 CTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 445
    || || || || || || || || || || || || || || || || || || || || ||
DB 513 GCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 572
QY 446 CGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 505
    || || || || || || || || || || || || || || || || || || || || ||
DB 573 GGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 632
QY 506 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 564
    || || || || || || || || || || || || || || || || || || || || ||
DB 633 GAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 692
QY 565 CTGCCACCGCGCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 624
    || || || || || || || || || || || || || || || || || || || || ||
DB 693 TGGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 752
```

RESULT 5

US-08-232-463-14/c
Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: SCHETTLINGER, F.

APPLICANT: DORNER, F.

TITLE OF INVENTION: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

QY 236 GGAGCCGGGGCCCCCTGCTGCTCTCCAGGCGGGAGCCCGCTTCCCGGCGG 295
11 1111 1111 11 111 111 111 111 111 111 111
Db 512 GGGACAGAGGGGGCCCCCGCCAGAGCCCCGGGGAGAGGGGGCGCCCGCGGCGG 453
QY 296 CTGGCGGGCAAGGGCGGGCTTGTACTCCCGGGGGCGGCCCCCGGCGAGAGA 355
11 1111 1111 11 111 111 111 111 111 111
Db 452 CTGCGCCCGGGGGGGCGGAGAGAAAGCAAGCGCCCGCGCGGGGCG-- 395
QY 356 AGAGACTGGGGGGGTTGGTCCCTGCTGCGCCCGGGGGCCACCAAGCAAGCCGG 415
11 1111 111 111 111 111 111 111 111 111
Db 394 ---CCGCGCGGGCGCATCCCCCCCCCTGTTTACCCCCCGGGGGGGCGCGG 338
QY 416 CATTGGGGGGAGACCTGCGCGAGCCGAGCGGCTGACAGCCCCCGGCGCAATATAGGC 475
11 1111 1111 11 111 111 111 111 111 111
Db 337 CCGCGGGGAGACGGGGGGCGGAGAGAGTTGCGGGGGCGCGCCGCAAAAAAATATTGCT 278
QY 476 GGTGCTGCCATTCAACAGGGGCTCTCTGTCGCGGGCGGCGCAAGAGCTTACGCTGGGC 535
11 1111 1111 111 111 111 111 111 111 111
Db 277 GGGCTAGCCCCGTAACCCCGGCGCTCGGGGCTCGGGGCTGCGCGCTGCGCGCG 218
QY 536 TGGGACAAGGGTGGGGGGCTCCCCGCTGCGCACCG--CTGGAACCGGGGGAGC 592
11 1111 111 111 111 111 111 111 111 111
Db 217 TGGCGGCGGGGGGGGCGCTCGCGCTCGCGCGGGCGCGGGGGCGCGCGCGCC 158
QY 593 CCCAGCACTACTCTGCGCGGGCGCGCAC--CTGCGCGCCACCGCCCGCGGAGCC 651
11 1111 1111 111 111 111 111 111 111 111
Db 157 CTCGCCGTCGCCCGCGCGGAGAGCGCGGGCGCGCTGCGCGCGGGGGGCG 98
QY 652 TGGCGGCGAGGAGGAGATGAGAGATGAGAGAGCCCTCTCGGGGGTGTGGCG 711
11 1111 111 111 111 111 111 111 111 111
Db 97 CGGCTCGCGCGGGGCTGCGCGGAGCGCGGAGGCGAGCGGGGGAGGGGCGG 38
QY 712 GCTCGG 717
11 111
Db 37 GCGCGG 32

RESULT 2
US-08-757-669A-16/c
; Sequence 16, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-757-669A-16

Query Match 1.6%; Score 70; DB 4; Length 801;
Best Local Similarity 50.8%; Pred. No. 5.9e-07;
Matches 247; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

QY 236 GGAGCCGGGGCCCCCTGCTGCTCTCCAGGCGGGAGCCCGCTTCCCGGCGG 295
11 1111 1111 11 111 111 111 111 111 111 111
Db 512 GGGACAGAGGGGGCCCCCGCCAGAGCCCCGGGGAGAGGGGGCGCCCGCGGCGG 453
QY 296 CTGGCGGGCAAGGGCGGGCTTGTACTCCCGGGGGCGGCCCCCGGCGAGAGA 355
11 1111 1111 11 111 111 111 111 111 111
Db 452 CTGCGCCCGGGGGGGCGCGGAGAGAAAGCAAGCGCCCGCGCGGGGCG-- 395
QY 356 AGAGACTGGGGGGTTCGCTGCTGCTGCGCGCGCGCGCGCGCAAGCAAGCGAG 415
11 1111 1111 11 111 111 111 111 111 111
Db 394 ---CCGCGCGGGCGCATCCCCCCCCCTGTTTACCCCCCGGGGGGGCGCGG 338
QY 416 CATTGGGGGGAGACCTGCGCGAGCCGAGCGGCTGACAGCCCCCGGCGCAATATAGGC 475
11 1111 1111 11 111 111 111 111 111 111
Db 337 CCGCGGGGAGACGGGGGGCGGAGAGAGTTGCGGGGGCGCGCCGCAAAAAAATATTGCT 278
QY 337 CCGCGGGGAGACGGGGGGCGGAGAGAGTTGCGGGGGCGCGCCGCAAAAAAATATTGCT 278
11 1111 1111 11 111 111 111 111 111 111
Db 476 GGTGCTGCCATTCAACAGGGGCTCTCTGTCGCGGGCGGCGCAAGAGCTTACGCTGGGC 535
11 1111 111 111 111 111 111 111 111 111
Db 277 GGGCTAGCCCCGTAACCCCGGCGCTCGGGGCTCGGGGCTGCGCGCTGCGCGCG 218
QY 536 TGGGACAAGGGTGGGGGGCTCCCCGCTGCGCACCG--CTGGAACCGGGGGAGC 592
11 1111 111 111 111 111 111 111 111 111
Db 217 TGGCGGCGGGGGGGGCGCTCGCGCTCGCGCGGGCGCGGGGGCGCGCGCGCC 158
QY 593 CCCAGCACTACTCTGCGCGGGCGCGCAC--CTGCGCGCCACCGCCCGCGGAGCC 651
11 1111 1111 111 111 111 111 111 111 111
Db 157 CTCGCCGTCGCCCGCGCGGAGAGCGCGGGCGCGCTGCGCGCGGGGGGCG 98
QY 652 TGGCGGCGAGGAGGAGATGAGAGATGAGAGAGCCCTCTCGGGGGTGTGGCG 711
11 1111 111 111 111 111 111 111 111 111
Db 97 CGGCTCGCGCGGGGCTGCGCGGAGCGCGGAGGCGAGCGGGGGAGGGGCGG 38
QY 712 GCTCGG 717
11 111
Db 37 GCGCGG 32

RESULT 3
US-09-230-371A-16/c
; Sequence 16, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 182 Seconds
(without alignments)
7366.982 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagtcctgcgaccccgacac.....aaaaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	70	1.6	801	2	US-08-770-379-16
C 2	70	1.6	801	4	US-08-757-669A-16
C 3	70	1.6	801	4	US-09-230-371A-16
C 4	67.2	1.5	936	1	US-08-018-977C-4
C 5	67	1.5	7218	1	US-08-232-463-14
C 6	62	1.4	801	4	US-09-298-568-3
C 7	61.2	1.4	4411529	4	US-09-103-840A-1
C 8	58.2	1.3	4403765	4	US-08-103-840A-2
C 9	58.2	1.3	4411529	4	US-09-103-840A-1
C 10	57.6	1.3	15872	4	US-09-105-537-1
C 11	57.6	1.3	43280	2	US-08-804-227C-1
C 12	56.6	1.3	8438	1	US-07-945-283-1
C 13	56.4	1.3	5852	1	US-07-867-106-2
C 14	56	1.3	4403765	4	US-09-103-840A-2
C 15	55.6	1.3	7898	4	US-08-984-709A-49
C 16	55.2	1.3	5539	4	US-08-628-829-3
C 17	54.6	1.2	7218	4	US-08-232-463-14
C 18	54.4	1.2	13987	2	US-08-804-227C-13
C 19	54.4	1.2	44377	2	US-08-804-227C-7
C 20	54.4	1.2	44377	2	US-08-804-198-1
C 21	54	1.2	1941	5	PCT-US91-07635-5
C 22	52.2	1.2	1325	4	US-09-412-102-3
C 23	52.2	1.2	1325	4	US-09-217-787-3
C 24	52.2	1.2	4257	2	US-08-690-473-1
C 25	52.2	1.2	4257	4	US-09-259-821A-1
C 26	52.2	1.2	4257	4	US-08-843-659-1
C 27	52.2	1.2	12001	1	US-08-458-568A-11

28	51.8	1.2	1227	1	US-08-924-254-1	Sequence 1, Appl
29	51.8	1.2	1227	1	US-08-924-254-3	Sequence 3, Appl
30	51.8	1.2	1227	3	US-09-120-249-1	Sequence 3, Appl
31	51.8	1.2	1227	3	US-09-120-249-3	Sequence 3, Appl
C 32	51.8	1.2	6453	1	US-08-306-691B-14	Sequence 14, Appl
C 33	51.8	1.2	6453	3	US-09-209-668-10	Sequence 10, Appl
C 34	51.8	1.2	6453	4	US-09-356-952-8	Sequence 8, Appl
C 35	51.6	1.2	5253	3	US-09-423-890-7	Sequence 7, Appl
36	50	1.1	3132	2	US-08-224-482-3	Sequence 3, Appl
37	50	1.1	3132	2	US-09-205-921-1	Sequence 2, Appl
38	49.8	1.1	1276	4	US-09-177-325-2	Sequence 2, Appl
39	49.8	1.1	1276	4	US-09-411-812A-2	Sequence 2, Appl
40	49.8	1.1	1276	4	US-09-590-113-2	Sequence 2, Appl
41	49.6	1.1	1225	1	US-08-286-020-1	Sequence 1, Appl
42	49.6	1.1	1225	1	US-08-603-919-1	Sequence 1, Appl
43	49.6	1.1	2186	2	US-08-878-546-9	Sequence 9, Appl
44	49.4	1.1	1241	1	US-07-593-657-6	Sequence 6, Appl
45	49.4	1.1	1241	4	US-08-942-012B-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-770-379-16/c
; Sequence 16, Application US/08770379
; Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770, 379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-770-379-16
Query Match 1.6%; Score 70; DB 2; Length 801;
Best Local Similarity 50.8%; Pred. No. 5.9e-07;
Matches 247; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

Db 245 GCCAGATTTCTGTTCTCCATCATTTTAATGCAACGTGTGGCTTTCCTACTCT 304
OY 3641 TACCATGTGATTTCTGTAACGTGTCCAAAAGAAAA 3682
Db 305 GACCAITGTAATGTTCTTAATAAATGTCCTCCAAAAGAAAA 346

RESULT 12
US-10-106-698-3799

Sequence 3799, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 3799
LENGTH: 196
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (7)..(8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (39)..(39)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (70)..(70)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (76)..(76)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (78)..(78)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (128)..(128)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (143)..(143)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (146)..(146)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (180)..(180)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (196)..(196)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-3799

Query Match 4.0%; Score 176.2; DB 9; Length 196;
Best Local Similarity 93.2%; Pred. No. 6.8e-37;
Matches 178; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3810 GGGGACTGCCAGTTTCTGACTGAGTAGAGTCCAGGAGTGTCTATCTGG 3869
Db 5 GGNNTTCCCGAGTTTCTGACTGAGTAGAGTCCAGGAGTGTCTATCCGG 64
OY 3870 CCAGATTAACTAGATTCTATTTCTGTTCTCTCCCTCCCTGAGACCTTATTTTA 3929
Db 65 CCAGATTAAAGNAATCTATTTCTGTTCTCTCCCTCCCTGAGACCTTATTTTA 124

OY 3930 TTGACCCCTCTTCTAGTTAAATTCCTTTGATTTGACTTTGTGAGAGAGGTTGGAC 3989
Db 125 TTGACCCCTCTTCTAGTTAAATTCCTTTGATTTGACTTTGTGAGAGAGGTTGGAC 184
OY 3990 AGTAGATTAGC 4000
Db 185 AGTAGATTAGC 195

RESULT 13

US-09-917-800A-737/C
Sequence 737, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendlick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 737
LENGTH: 565
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1102820
US-09-917-800A-737

Query Match 2.5%; Score 107.2; DB 10; Length 565;
Best Local Similarity 62.6%; Pred. No. 5.3e-18;
Matches 259; Conservative 0; Mismatches 133; Indels 22; Gaps 5;

OY 2512 CTGTGATACAGTCATGTGGGAAGGATTTGGCTGTGATTTTTCACGTAAATGAT 2571
Db 422 CTGGACAGTGTGCAATGTGGGAAGACATGTGGCTG-CCTTACTTCTTAACTTATGAGT 364
OY 2572 AACAAATTTCTTACTGCTCAAAAACCAAAATCTTTGAAAAGAGTGGGATGTTAGT 2631
Db 363 AACGTAGTTCTTCCACCCCAACAGAGAAATGTTAGAAAAGGGTTGGGGTTAG 304
OY 2632 TTCAGAACACAGTTACAGCTGTAAACAAAAGCACTTAGTATTGGGATGTCACCAAAAC 2691
Db 303 TTCCTG-----AAGGATACAGTTGCTCACTTAGGATGTCACACACAGGT 260
OY 2692 CTGTATAATGTCTCTGTATGATCACTTTCTCAAGTATTCCTTCATTTGGGCTTCATCCT 2751
Db 259 CTGTGAGATGTCTCTGTGCAATGTCACTTCTCAAGTATTCCTTTGGGGCTTCAGCCT 200

QY 3867 TGG 3869
Db 539 TTG 541

RESULT 7

US-09-918-995-34149
; Sequence 34149, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34149
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34149

Query Match 8.9%; Score 389.6; DB 9; Length 411;
Best Local Similarity 98.8%; Pred. No. 6.4e-94;
Matches 403; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 2179 GGCAGATTGAGAACTTTCAGGGGCTCTTTCTATACATTTGCTATGTTAAAGGGGTAAA 2238
Db 1 GGCAGATTGAGAACTTTCAGGGGCTCTTTCTATACATTTGCTATGTTAAAGGGGTAAA 60
QY 2239 AGGCGCTCTTCATATACATGTGGAAGATGAAACACCCCTTCTTAAAGCTGTGCT 2238
Db 61 AGGCGCTCTTCATATACATGTGGAAGATGAAACACCCCTTCTTAAAGCTGTGCT 120
QY 2299 GCATGCGACTTCTTCACCCCTGGTACACCCCTTATAGTGGATATGATTTTAAAC 2358
Db 121 GCATGCGACTTCTTCACCCCTGGTACACCCCTTATAGTGGATATGATTTTAAAC 180
QY 2359 CTAATAATTAACAACAACCTCACCATG-ACCTTTAGGACAGAGGAATGCAAGTG 2417
Db 181 CTAATAATTAACAACAACCTCACCATGAGCTTTAGGACAGAGGAATGCAAGTG 240
QY 2418 AAGGATGAGACGACCATCTTCACAGATAGAAAGACATGGAGAGTTGCTAGTAAC 2477
Db 241 AAGGATGAGACGACCATCTTCACAGATAGAAAGACATGGAGAGTTGCTAGTAAC 300
QY 2478 TGTCTGAAAAGATAGTGTTCATTTGAAACTATTCTGTGATACAGTCATGTGGAGGGA 2537
Db 301 TGTCTGAAAAGATAGTGTTCATTTGAAACTATTCTGTGATACAGTCATGTGGAGGGA 360
QY 2538 TGTTCGCTGTGATATTTTTCAGTTAATGATACAAATTTCTTAC 2585
Db 361 TGTTCGCTGTGATATTTTTCAGTTAATGATACAAATTTCTTAC 408

RESULT 8

US-09-918-995-30064
; Sequence 30064, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30064
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30064

Query Match 8.2%; Score 357; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 3.3e-85;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 919 AGCTGTTTCAGGCTATGAACTGAGAGAGAGAGAAACAGACTATCTGAGAAATTA 978
Db 9 AGCTGTTTCAGGCTATGAACTGAGAGAGAGAGAAACAGACTATCTGAGAAATTA 68
QY 979 AACTGAGAGTGCACCCGAGCTTCCGAGACATCCCACTCTCCCAAGCCCTCT 1038
Db 69 AACTGAGAGTGCACCCGAGCTTCCGAGACATCCCACTCTCCCAAGCCCTCT 128
QY 1039 CATGTGGCGGAGTGAAGAGACATTAAGAAATCCCATTTGAAATGACAGAA 1098
Db 129 CATGTGGCGGAGTGAAGAGACATTAAGAAATCCCATTTGAAATGACAGAA 188
QY 1099 AGACTCTGTTAAAAAGCTGCTCTGAATTTTCAAAAGTCAAAAACAACTCTAAC 1158
Db 189 AGACTCTGTTAAAAAGCTGCTCTGAATTTTCAAAAGTCAAAAACAACTCTAAC 248
QY 1159 ACTCTCTATTAAGAGAACCCCTGTGTCTTATCTGAACAGTGTGTTAAAGTGAAT 1218
Db 249 ACTCTCTATTAAGAGAACCCCTGTGTCTTATCTGAACAGTGTGTTAAAGTGAAT 308
QY 1219 TGAGAGCCCAAGAAACCCCAAGAAAGCCCGTCTTCAAGTGAACCCCAAGAC 1275
Db 309 TGAGAGCCCAAGAAACCCCAAGAAAGCCCGTCTTCAAGTGAACCCCAAGAC 365

RESULT 9

US-09-917-800A-16/c
; Sequence 16, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 590

Query	Match	Similarity	Score	51.6	DB	10	Length	615
<p>Best Local Similarity 94.6% Pred. No. 2.9e-137</p> <p>Matches 582; Conservative 0; Mismatches 30; Indels 3; Gaps 1</p>								
QY	112	CGGCCCCCTGCGGGGGGCAATCCTTGAGCAGCACTGAGTACGAGCGGGCTGGCGGCTGG	171					
Db	1	CGGCCCCCTGCGGGGGGCAATCCTTGAGCAGCACTGAGTACGAGCGGGCTGGCGGCTGG	60					
QY	172	CGGGGGCCCGAGGAGAGAGCTTGGGGCGGGCCGAAGCCACTTCTCTCCCGGACCGTAAG	231					
Db	61	CGGGGGCCCGAGGAGAGAGCTTGGGGCGGGCCGAAGCCACTTCTCTCCCGGAGCTAAG	120					
QY	232	TCAAGAGAGCCGGGGGCCCCCGCTGGAGCTCTCCAGAGGGGGAGCCCGCCGCTTCCCGG	291					
Db	121	TCAAGAGAGCCGGGGGCCCCCGCTGGAGCTCTCCAGAGGGGGAGCCCGCCGCTTCCAG	180					
QY	292	CGGCT--GCGGCGGCAAGGGGCGGGGCTTGTACTCCCGGCGGGGGCGCCCGGCG	348					
Db	181	CTGGCTGGGGGCGGCGGCGGCGGCGGCTTGTACTCCCGGCGGGGGCGCCCGGCG	240					
QY	349	AGCAGGAAGAGAGCTGGGGGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	408					
Db	241	AGCAGGAAGAGAGCTGGGGGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	300					
QY	409	AAGCGGCAATTTGGGGGGGAGGCTTGCAGCGGAGCCGAGCCGAGCCGCGCCCAAGT	468					
Db	301	AAGCGGCAATTTGGGGGGGAGGCTTGCAGCGGAGCCGAGCCGAGCCGCGCCCAAGT	360					
QY	469	ATCAGGCGGCTCTCCCTTTCAGAGAGGCTCTCTCTGCTGGGGCGGGCCCAAGAGCTTACG	528					
Db	361	ATCAGGCGGCTCTCCCTTTCAGAGAGGCTCTCTCTGCTGGGGCGGGCCCAAGAGCTTACG	420					
QY	529	CCTGGGCTGGGAGAAAGGAGTGGGGCGGCTCCCGGCTGCGCACCGCTCGAGCCGGGG	588					
Db	421	CCTGGGCTGGGAGAAAGGAGTGGGGCGGCTCCCGGCTGCGCACCGCTCGAGCCGGGG	480					
QY	589	GACCCCAACACTTCTGCCCCGGGGCGCAACCCCTTGCGGCCACCGGCACCGCGGGA	648					
Db	481	GACCCCAACACTTCTGCCCCGGGGCGCAACCCCTTGCGGCCACCGGCACCGCGGGA	540					
QY	649	CCCGGGGGGGGAGGAGGAGATGAGAGATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAG	708					
Db	541	CCCGGGGGGGGAGGAGGAGATGAGAGATGAGAGAGATGAGAGAGAGGAGGAGGAGGAG	600					
QY	709	GCGGCTCGGAGAGCT 723						
Db	601	GCGGCTCGGAGAGCT 615						

QY	617	GCCACCCCTCGGCCACC GCCACC GCGGAGCCTGGCGGC CAGCGGGCAGATGAA	678
Db	2	GCCACCCCTCGGCCACC GCCACC GCGGAGCCTGGCGGC CAGCGGGCAGATGAA	61
QY	677	GAGTATGAGGAGAGAGCCCTCTCGGGG GTGTGGCGCTCGGAGGCTTCAGTCAGGCCG	736
Db	62	GAGTATGAGGAGAGAGCCCTCTCGGGG GTGTGGCGCTCGGAGGCTTCAGTCAGGCCG	122
QY	737	CTTGCTCAAAACAGATCTTTCTGCTCAATTGGACCTCATCGAACGACGACGACGCT	796
Db	122	CTTGCTCAAAACAGATCTTTCTGCTCAATTGGACCTCATCGAACGACGACGACGCT	181
QY	797	GCAGGCGCAAGAAAAGAGATGCAGAGAGCTGAAGTCAGAGAGACAGCAGCTCTTGGCTCG	856
Db	182	GCAGGCGCAAGAAAAGAGATGCAGAGAGCTGAAGTCAGAGAGACAGCAGCTCTTGGCTCG	241

Db 346 -HOKOSRITPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRL 536
Db 402 --TELKARLRLNFOROLRQ-EVYVCMRDOTALFETALNAKAYKRSKQSLREARITTEK 458
QY 537 RM---YKKKGIOSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQIEYLSIIQHARD 485

RESULT 13
US-09-535-008-69
; Sequence 69, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1678
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-69

Query Match 5.3%; Score 171; DB 4; Length 1678;
Best Local Similarity 19.9%; Pred. No. 0.0023;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKEBPPLASSOGSPAPSPAGCGGKGRGLL--PAGAARQOEESWGSVPLPCPPA- 108
Db 1 MSTPDPPL-----GSTRPGRSPGPGSPGAMLGPSGPGSHSMWG---PSGPPSAG 53
QY 109 --TKQAGIGEP-----AAAGACSPRPXYQAV--LPITGSLVAAKERTPW 152
Db 54 HPITPGPGGYPODNNHQMHRPMESMHEKMSDDPRYNQMGKMGSRGAGMGPPSPM 113
QY 153 AGDKGAASPAATASDPAGPPLPLPGPPLATATAGTIASEGKRSKRSKPLGGGG 212
Db 114 DQHSQGYPSPLGSEHASSVPASGSSG-----QMSGP-GGAPL 154
QY 213 SGASSQAACIKQILLDLIEQOOOOLAKEKEIEELKSE--RDTLLARIEEMERMOL 270
Db 155 DGADPQA-----LGOONRGPTPENQMLHQLRAQIMAYKMLAGPLPHLQW 202
QY 271 VKKDNEKRRHKLFGYETERETELSEKIKLECOBELSTQTLPPKPRSCGSGK-- 328
Db 203 A-----VOGKRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPGSTERKTPYKKLAPESKYVTKPKHSPIKEEPCGSLSET 375
Db 236 PGCGPGGAPRPRYSRPHMGGMNPPRGSSGVRPGMPGQ---PGGPRKPMPEGMANA 292
QY 376 VCKRELRSQETPRKPRSSVDTPLRSTPOKGPSTHREKAFSSIEDLPYLSTTEMYL 435
Db 293 AAPTSTPQKILIPQIPGRSPAPPAVPAPSPMPOTGSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVAVCL-----MPSSVAGETSVLAVPSW 478
Db 346 -HOKOSRITPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRL 536
Db 402 --TELKARLRLNFOROLRQ-EVYVCMRDOTALFETALNAKAYKRSKQSLREARITTEK 458

QY 537 RM---YKKKGIOSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQIEYLSIIQHARD 485

RESULT 14
US-09-535-008-65
; Sequence 65, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 1679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-65

Query Match 5.3%; Score 171; DB 4; Length 1679;
Best Local Similarity 19.9%; Pred. No. 0.0023;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKEBPPLASSOGSPAPSPAGCGGKGRGLL--PAGAARQOEESWGSVPLPCPPA- 108
Db 1 MSTPDPPL-----GSTRPGRSPGPGSPGAMLGPSGPGSHSMWG---PSGPPSAG 53
QY 109 --TKQAGIGEP-----AAAGACSPRPXYQAV--LPITGSLVAAKERTPW 152
Db 54 HPITPGPGGYPODNNHQMHRPMESMHEKMSDDPRYNQMGKMGSRGAGMGPPSPM 113
QY 153 AGDKGAASPAATASDPAGPPLPLPGPPLATATAGTIASEGKRSKRSKPLGGGG 212
Db 114 DQHSQGYPSPLGSEHASSVPASGSSG-----QMSGP-GGAPL 154
QY 213 SGASSQAACIKQILLDLIEQOOOOLAKEKEIEELKSE--RDTLLARIEEMERMOL 270
Db 155 DGADPQA-----LGOONRGPTPENQMLHQLRAQIMAYKMLAGPLPHLQW 202
QY 271 VKKDNEKRRHKLFGYETERETELSEKIKLECOBELSTQTLPPKPRSCGSGK-- 328
Db 203 A-----VOGKRP-----MGMOQOMPTLPPPSVATGPGPG 235
QY 329 -----HKRSPGSTERKTPYKKLAPESKYVTKPKHSPIKEEPCGSLSET 375
Db 236 PGCGPGGAPRPRYSRPHMGGMNPPRGSSGVRPGMPGQ---PGGPRKPMPEGMANA 292
QY 376 VCKRELRSQETPRKPRSSVDTPLRSTPOKGPSTHREKAFSSIEDLPYLSTTEMYL 435
Db 293 AAPTSTPQKILIPQIPGRSPAPPAVPAPSPMPOTGSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVAVCL-----MPSSVAGETSVLAVPSW 478
Db 346 -HOKOSRITPIQKRGDLPVEILQEREYRLQARIARHIOLENLPGSLADLTAKT--- 401
QY 479 RDHSEVLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQRIREORILQRL 536
Db 402 --TELKARLRLNFOROLRQ-EVYVCMRDOTALFETALNAKAYKRSKQSLREARITTEK 458
QY 537 RM---YKKKGIOSEPEVTSFPEPDD 560
Db 459 QOKIEQERKRKRKHQIEYLSIIQHARD 485

```

OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTTPKHSPIKEEPCGSLSET 375
DB 236 PGPGRPGAPAPNYSRPHGMGPMPPRPGSVPPMGQ---PPGPRPKPMEEGMANA 292
OY 376 VCKRELISOETPEKPRSSVDTPRLSTPOKSPSTHREKAFSSIEDLYLSTTEMYLOR 435
DB 293 AAPSTPOKLIIPQPTGRPSAPPAVPAPASVPMPTOTOSPOQAPAPMVL----- 345
OY 436 WHOPPSPLP-----LRSSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
DB 346 -HOKOSRITPIQPRGLDPEVILQEREYRLQARIHRIQLEENLPGSLADLRTKAT--- 401
OY 479 RDSVLEPLDPPNSDLLENLDSVFSKRAKLE--LDEKRRKMDIQIRIQRIQRIQLOL 536
DB 402 --TELKALRLNFOQLRO-EVYVCMRDTALETNALNAKAYKRSKROSLREARITTEKLE 458
OY 537 RM---YKKGIOESEPEVTSEFPEDD 560
DB 459 OOKIEORRRKQKHQEYLNSTILOHAKD 485

```

```

RESULT 11
US-09-535-008-75
; Sequence 75, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 75
; LENGTH: 1649
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-75

```

```

Query Match
Best Local Similarity 5.3%; Score 171; DB 4; Length 1649;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

OY 52 LKEPPPLASSQGGSPAPSPAGCGKGRGLL--PAGAARQOEESMGSGVPLPCPPA- 108
DB 1 MSTDPPL-----GSTRPGSPGPGSPGAMLGSPGSPGSAHSMKG--PSPGPPSAG 53
OY 109 --TKOAGIGBP-----AAAGACSPRKYQAV--LPITGSLVAAAKEPTPW 152
DB 54 HPIPTQGRGGYPODNHOMHKPMESMHEKMSDDPRYNQMKGMGSGHAGMGPPSPM 113
OY 153 AGDGGAAPATAADAPAPPLPLPDPPLAATAAGTAAASEGKRWKSMKSPGLGGGG 212
DB 114 DQHSQGTSPSLGSEHASSVPASGSSGP-----QMSGGP--GGAPL 154
OY 213 SGASSQAACIKOTILLDLIEQOOQOLAKEREIEELKSE--RDTLLARIEMRMQOL 270
DB 155 DGADPQA-----LGOQNRGPTFPNNOHOLHQAQIMAYKMLARQPLPDHLOM 202
OY 271 VKKDNEKERHKLFOGYETEEREETELSEKIKLEQOPELSETSQLTPKPPSCGSGK- 328
DB 203 A-----VOGRKP-----MPGMOOQMPILPPLPSVATGPGPG 325
OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTTPKHSPIKEEPCGSLSET 375
DB 236 PGPGRPGAPAPNYSRPHGMGPMPPRPGSVPPMGQ---PPGPRPKPMEEGMANA 292
OY 376 VCKRELISOETPEKPRSSVDTPRLSTPOKSPSTHREKAFSSIEDLYLSTTEMYLOR 435

```

```

RESULT 12
US-09-535-008-71
; Sequence 71, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 71
; LENGTH: 1650
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-71

```

```

Query Match
Best Local Similarity 5.3%; Score 171; DB 4; Length 1650;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

OY 52 LKEPPPLASSQGGSPAPSPAGCGKGRGLL--PAGAARQOEESMGSGVPLPCPPA- 108
DB 1 MSTDPPL-----GSTRPGSPGPGSPGAMLGSPGSPGSAHSMKG--PSPGPPSAG 53
OY 109 --TKOAGIGBP-----AAAGACSPRKYQAV--LPITGSLVAAAKEPTPW 152
DB 54 HPIPTQGRGGYPODNHOMHKPMESMHEKMSDDPRYNQMKGMGSGHAGMGPPSPM 113
OY 153 AGDGGAAPATAADAPAPPLPLPDPPLAATAAGTAAASEGKRWKSMKSPGLGGGG 212
DB 114 DQHSQGTSPSLGSEHASSVPASGSSGP-----QMSGGP--GGAPL 154
OY 213 SGASSQAACIKOTILLDLIEQOOQOLAKEREIEELKSE--RDTLLARIEMRMQOL 270
DB 155 DGADPQA-----LGOQNRGPTFPNNOHOLHQAQIMAYKMLARQPLPDHLOM 202
OY 271 VKKDNEKERHKLFOGYETEEREETELSEKIKLEQOPELSETSQLTPKPPSCGSGK- 328
DB 203 A-----VOGRKP-----MPGMOOQMPILPPLPSVATGPGPG 325
OY 329 -----HKRSPFGSTERKTPVKKLAPESKVTKTTPKHSPIKEEPCGSLSET 375
DB 236 PGPGRPGAPAPNYSRPHGMGPMPPRPGSVPPMGQ---PPGPRPKPMEEGMANA 292
OY 376 VCKRELISOETPEKPRSSVDTPRLSTPOKSPSTHREKAFSSIEDLYLSTTEMYLOR 435
DB 293 AAPSTPOKLIIPQPTGRPSAPPAVPAPASVPMPTOTOSPOQAPAPMVL----- 345
OY 436 WHOPPSPLP-----LRSSPKKEETVARCL-----MPSVAGETSVLAVPSW 478

```

Db 114 DQSGCYPSPLGSGEHSASPVPASGSPSGP-----OMSSGP--GCAPL 154
QY 213 SGASSQAACLKQILLDLIEQOOQOLAKEKEIEELKSE--RDTLLARIEMERRMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNNQHLQRAQIMAYMLARGOPLPDHLOM 202
QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGKG-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVSATGPGPG 235
QY 329 -----HKRSPFGSTERKTPYKLAPEFSKYKTKPKHSPIKEEPCGSLSET 375
Db 236 PGPGPGPAPAPVNSPHMGGNMPPPGSPGVPCMPGQ---PCGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPEKPSVSDTPPRLSTPOKGSTHKEKAFSEIEDLPLSTTEMYLOR 435
Db 293 AAPTSTPQKLIIPQPIGRSPAPAPVPAAPVMPQOTSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
Db 346 -HQKSRITPIQKPRGLDPEVILQEREYRLQARIARIQLELENLPGSLAGDLTKAT--- 401
QY 479 RDHVSPLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQIRIQRILLORLOL 536
Db 402 --TELKALRLNFQRLRO-EVYVCMRBDTALETALNAKAYKRSKRSILREARITEKLEK 458
QY 537 RM---YKKKGIOESEPEVTSFPEPDD 560
Db 459 QOKIEQERRKROKHQEYLSILOHARD 485

RESULT 9

US-09-535-008-67
; Sequence 67, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 1646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-67

Query Match 5.3%; Score 171; DB 4; Length 1646;

Best Local Similarity 19.9%; Pred. No. 0.0022;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKRPBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARPGQOEESVGLPCPPPA- 108
Db 1 MSTPDDPL---GTRPRGSPGPGSPGAMLGSPSPGSAHSMWG---PSPGPPSAG 53
QY 109 --TKAGIGGEP-----AAAGAGCSRPXYQAV--LPITGSLVAIAAKRPTPW 152
Db 54 HPIPTGPGGYPDNDNHQNHKPMESNHEKMSDPRYNQMKGMGSGHAGGPPSPM 113
QY 153 AGDGAASPAATASDPAGRPPLPLPGPPLAPATAGTLAEGRWKSMRKSPLGCGG 212
Db 114 DQSGCYPSPLGSGEHSASPVPASGSPSGP-----OMSSGP--GCAPL 154
QY 213 SGASSQAACLKQILLDLIEQOOQOLAKEKEIEELKSE--RDTLLARIEMERRMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNNQHLQRAQIMAYMLARGOPLPDHLOM 202

QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGKG-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVSATGPGPG 235
QY 329 -----HKRSPFGSTERKTPYKLAPEFSKYKTKPKHSPIKEEPCGSLSET 375
Db 236 PGPGPGPAPAPVNSPHMGGNMPPPGSPGVPCMPGQ---PCGPPKPMPEGMANA 292
QY 376 VCKRELRSOETPEKPSVSDTPPRLSTPOKGSTHKEKAFSEIEDLPLSTTEMYLOR 435
Db 293 AAPTSTPQKLIIPQPIGRSPAPAPVPAAPVMPQOTSPGQAPAPAPVPL----- 345
QY 436 WHOPPSPLP-----LRESSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
Db 346 -HQKSRITPIQKPRGLDPEVILQEREYRLQARIARIQLELENLPGSLAGDLTKAT--- 401
QY 479 RDHVSPLRDPNPSDLLENLDSVFSKRAKLE--LDEKRRKMDIQIRIQRILLORLOL 536
Db 402 --TELKALRLNFQRLRO-EVYVCMRBDTALETALNAKAYKRSKRSILREARITEKLEK 458
QY 537 RM---YKKKGIOESEPEVTSFPEPDD 560
Db 459 QOKIEQERRKROKHQEYLSILOHARD 485

RESULT 10

US-09-535-008-2
; Sequence 2, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-2

Query Match 5.3%; Score 171; DB 4; Length 1647;

Best Local Similarity 19.9%; Pred. No. 0.0022;

Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

QY 52 LKRPBPPLASSOGSPAPSPAGCGGKRGILL--PAGAARPGQOEESVGLPCPPPA- 108
Db 1 MSTPDDPL---GTRPRGSPGPGSPGAMLGSPSPGSAHSMWG---PSPGPPSAG 53
QY 109 --TKAGIGGEP-----AAAGAGCSRPXYQAV--LPITGSLVAIAAKRPTPW 152
Db 54 HPIPTGPGGYPDNDNHQNHKPMESNHEKMSDPRYNQMKGMGSGHAGGPPSPM 113
QY 153 AGDGAASPAATASDPAGRPPLPLPGPPLAPATAGTLAEGRWKSMRKSPLGCGG 212
Db 114 DQSGCYPSPLGSGEHSASPVPASGSPSGP-----OMSSGP--GCAPL 154
QY 213 SGASSQAACLKQILLDLIEQOOQOLAKEKEIEELKSE--RDTLLARIEMERRMOL 270
Db 155 DGADPOA-----LGOONRGPTFPNNQHLQRAQIMAYMLARGOPLPDHLOM 202
QY 271 VKKDNEKERNKLEFGYETEERETELSEKIKLECOPELSETQTLPPKPFSGRSGKG-- 328
Db 203 A-----VQGRP-----MGMOQOMPTLPPPSVSATGPGPG 235

```

QY 159 AASPAATASDPAGPPPLPGLPPPLPATATAGTLAASEGRW-----KSMKSPILGGGG-- 213
DB 788 AEPPLATEAGTGPRLPLPSV--SPSOGAPILPSSEASAPAPDALPDSPPATGGEV 845
QY 214 GASSQAACLOKILLDLEQOQOOLAKKEIEE-----LKSPEDTL 257
DB 846 SAILASALNG-----SSSPSEVEAPSSSEDEPTAEATSGIFDTSSDGLQARPPV 896
QY 258 LARLERMERRM-----OLVKRD-----NEKEHKLPGYTEEREE 293
DB 897 VPAFRSLQKQVGPDDSLDIDIPSSADGGEVFPSSATGPGSGQPRALDSGYDENYES 956
QY 294 TELSEKIKLE-CQPE-----LSETSOTLPPKPPSCGSGKHRR-----KSPGSPERKT- 342
DB 957 PEVFLKAOGCEPOAFELASEGEGPPETRLSTLSGLNKNPVRDAYSIDLEAEAE 1016
QY 343 ----PVKKL-----APFSKYTKTPKHSPIKEEPCGLSETVCVKRELRSOETPEKPRS 392
DB 1017 ATSGPPEKCGGDAPPELGLPSTGCP-----SEQVCLRGVSGEAGSGSGPG 1063
QY 393 SVTTPPRL-----STPKGPST-----HPKKAFFSETIEDLPYLSTTEMYLCRNHOPPP 441
DB 1064 EV-LPPLQLDESSPE--PSTCPGLVPEPPEPGAPAVRPGPSPSCQFLL----- 1113
QY 442 SPLPLESSPKKEETVARCLMSSVAGETSVLAVPMDH-----SVEPLR----- 487
DB 1114 TPVPLRSEGNSSD-----FQGPGLSGPAPQKRMGPGTPAPRRLALPLPA 1162
QY 488 ---DPNSSLLENLDDSVFSKRAKLELDEKRRKMDIR-----IREORT 530
DB 1163 ALEGREEDSEEDSED-----ESDELR-CISVQPEDESEAPAVPVVAESQS 1213
QY 531 IQRLOLMYKKKGIOE-----EPEVTSFPEPDDVESLITPFLPVAFGRPLP--K 581
DB 1214 ABNLISLKMBSLSETCEDLERKKAVSFF---DVTYTLFQESPTRELGLPFFGAK 1270
QY 582 LTPQNF 587
DB 1271 ESPPTE 1276

RESULT 7
US-09-535-008-63
; Sequence 63, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-63

Query Match
Best Local Similarity 19.9%; Score 171; DB 4; Length 543;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

QY 52 LKEGPPPLASSOGSPAPSPAGCGGKGLLL--PAGAAPGOEESWGSGVPLPCPPPA- 108
DB 1 MSTPDPPL-----GGTPRPGSPGPGSPGAMLGPSGSPSASMSMG---PSPGPPSAG 53
QY 109 --TKAGIGGP-----AAAGAGCSPPRKYQAV--LPITQGSILVAAKEPTPW 152

```

```

DB 54 HPITPGGGGIFPDNMHMKPMESNHEKMSDDPRYNOKMGMRSGHAGMPPSPM 113
QY 153 AGDKGAASPAATASDPAGPPPLPGLPPPLPATATAGTLAASEGRWKSMSKSPILGGGG 212
DB 114 DOHSQGSPLSGSEHSAHSPVAPGSPSSCP-----QMSGP--GGAPL 154
QY 213 SGASSQAACLOKILLDLEQOQOOLAKKEIEELASE--RDTLLARIERMERROL 270
DB 155 DGADPQA-----LGQONRGPTPNQOLQRLQIAVYKMLARGOPLPDHLM 202
QY 271 VKKONERKHLFGYTEERETELSEKIKLECOPELSTOTLPPKPPSCGSGKG-- 328
DB 203 A-----VOGRKP-----MGMOQOMTTLPPSVSATGPGPG 235
QY 329 -----HKRSPGSTERKTPVAKLAPEFSKYTKTPKHSPIKEEPCGLSET 375
DB 236 PGPGPGPAPPNISRPHGGMGPMPPGSGVPPGMPQ---PFGPPKPMPEGPMANA 292
QY 376 VKRELRSOETPEKPPSSVDTPPRLSTPKGPSTHPKKAFFSEIEDLPYLSTTEMYLCR 435
DB 293 AAPTSTQKILPPQPRGSPAPPAVPAASVMPPTQSPGQAPAPMVP----- 345
QY 436 WHPPSPPL-----LRESSPKKEETVAKCL-----MPSVAGETSVLAVPMDH 478
DB 346 -HOKOSRITPQKRGDLDEVEILOERETRLQARIARIQLENLGLSLAGDLRTAT--- 401
QY 479 RDSVPELDPNPSDLLENLDDSVFSKRAKLE--LDEKRRKMDIRQIRQILQRL 536
DB 402 --TELKALILNPNQRLQ--EYVVCMKRDTALETALNKAAYRSKRSKRSIAREARTKLEK 485
QY 537 RM---YKKKGIOSEPEVTSFPEPDD 560
DB 459 QOKIEQERRKROKHQEYLSILQHAKD 485

RESULT 8
US-09-535-008-61
; Sequence 61, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; EARLIER FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-61

Query Match
Best Local Similarity 19.9%; Score 171; DB 4; Length 577;
Matches 113; Conservative 65; Mismatches 249; Indels 140; Gaps 21;

```

```

QY 52 LKEGPPPLASSOGSPAPSPAGCGGKGLLL--PAGAAPGOEESWGSGVPLPCPPPA- 108
DB 1 MSTPDPPL-----GGTPRPGSPGPGSPGAMLGPSGSPSASMSMG---PSPGPPSAG 53
QY 109 --TKAGIGGP-----AAAGAGCSPPRKYQAV--LPITQGSILVAAKEPTPW 152
DB 54 HPITPGPGGYPDNMHMKPMESNHEKMSDDPRYNOKMGMRSGHAGMPPSPM 113
QY 153 AGDKGAASPAATASDPAGPPPLPGLPPPLPATATAGTLAASEGRWKSMSKSPILGGGG 212

```

QY 487 RDNPSD 493
Db 1126 KPPPTPE 1132

RESULT 5
US-09-370-838-216
; Sequence 216, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Monamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ. ID NOS: 289
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-216

Query Match 5.4%; Score 174; DB 4; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.00034;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA---ALGCPDEPDAE-AHFLPHRRLK-----EGPPLASOGSGSPAPSP 71
Db 8 EKAENALGSPPEERKVLNENGLTPPRREKALENGLSPPEAGEKVLVNGGLTPPKSE 67
QY 72 AGCGGKRGILLPAGA-----APGOE---ESMGGSVPL--PCP-----PP 107
Db 68 DKVSENG-GLRFPNTRPPTGWRAPGPWEKTPESMGAPITIGEPAPETSLERAPPS 126
QY 108 ATQAGIGEPAAAGACSPRYOAVLP-----IOTGSLVAAKEPTPWAG--DKG-G 158
Db 127 AVVSSRNGGETAGPLG--PAPKNGTLEPTERRAPETGAPRA---PGAGRLDUGSG 179
QY 159 AASPAATADPAG-----PPPLLPGRP----- 181
Db 180 GRAVGTGTAPGGGGGVADAKAGWVNTTRPPPPPLPPPEAPQPRLEPPAPRAPREV 239
QY 182 ---PLAPTATAGTLAASGRW---KSMRKSPLGGGSGGASQAACLQILLLODLI 233
Db 240 APESEPGAPDSRAGDGTALSGDGPPEKRGK-----EMRPLFLDLG 282
QY 234 EQOOOOLAKEKELEKSERDILLARIEMERMOLVKDNEKERHKLFOGYET----- 288
Db 283 PPGNSFOIKAR-----LSRLSLALPLTLTPPGPGPRRRPMEGADAGAAG 330
QY 289 -----ERRETELSEKIKLECPBELSTQTLPPKPSGCRSGKHRRKSPF- 335
Db 331 EAGGAGAPPAEEDGEDEDEDE-----EDEEAAPGA-AAAPRGGRARRAIVP 379
QY 336 ---GSTERKTPYKLAPEFSKVTPTPKHSPIKEPSCSLSETVCKRELRS----- 383
Db 380 VVVSASADADARPLRGL-----LKSFGADPED-----SLEKKRKNVSHGQYV 426
QY 384 ---QETPEK-----PRSSVD-----TPRLSTPQK--PSTHPKKAFFSSEI--- 420
Db 427 YLFDOETPTNELSVQAPPEGDPTDPSTPPAPPTPPHAPTPGDFPS---NDSGFGSFEWA 483
QY 421 EDLPYLSTEMYLCRMHQPPSPPLPLRESSPKKEETVACRLMPSVAGESVLAIVPMRD 480
Db 484 EDEPLT-----PPGP-PL-----CFSRSV---SPALETGP--- 511

QY 481 HSYEPLRDPN--PSDLLEN 497
Db 512 ---PPRAPDARAPACPVEN 527

RESULT 6
US-08-976-255-11
; Sequence 11, Application US/08976255
; Patent No. 6136581
; GENERAL INFORMATION:
; APPLICANT: Jono, Keith E.
; APPLICANT: Plozman, Gregory
; TITLE OF INVENTION: KINASE GENES AND USES
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; OPERATING SYSTEM: IBM compatible
; SOFTWARE: FASTSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,255
; FILING DATE: No. 6136581ember 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031,675
; FILING DATE: No. 6136581ember 22, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 229/182
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; ;
; ;
; INFORMATION FOR SEQ. ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1384 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-976-255-11

Query Match 5.3%; Score 173; DB 4; Length 1384;

Best Local Similarity 20.9%; Pred. No. 0.0013;
Matches 164; Conservative 79; Mismatches 265; Indels 278; Gaps 42;

QY 10 AAAAPAGGNPEGRUDYERAAAL-----GCPDEPDAE-AHFLPHRRLKRP----- 55
Db 561 APSPRATADDDDDSDGTAASLAMEPLLGHGPPYDVPMGNGDHY--FRSLARPPLCPSPS 619
QY 56 ---GPPLASSOGGS-----PA-----PSPACCG-----KGRGLLPAGA 87
Db 620 PSPSAGTSLAEGGADADWGAFCAPFEDPLGTSPLGSSGAPPLPLTGBDELBEVGA 679
QY 88 APGOEESWGGSVPL-----PCPPATKQAGIGGEPAAGA--GC-----SPRP 129
Db 680 RRAAQRGHRMNSVANNSSRCPE-----SMDPVSAAGHAECGCPSPKQTPRASPE 731
QY 130 KY--QAVLPLOTGSLVAAKEP-----TP-W-----AGDKG 158
Db 732 GTFEPBLGLQ-----AASAOEPCCEPLPHLCSAOGIAPAPCLVITSWTETASSGGDHPQ 787


```

Db      463 VPSEPTSTTANLGLLSRPSVCPRLPLGPENHRAGSNEDPLAPSGTPPTIPDETFF 522
QY      127 ---PRPKY-----QAVLPLOTGSLVAANE-----PTPMAGDKG 157
Db      523 GGRVPRPAVHYHKEASDVEISLSDSDSVIYVEGLPLPPPPSGATPPPIAPTGP 582
QY      158 GAASPAATASD-----PAGPPPLPLPGPP---LAPTATAGTLAASGKMKSPPLG 209
Db      583 PTASPVPAKKEPEELPAAPGLPPPPPPPPVGVXLLPPQLVPEG-----TP--G 633
QY      210 GGGSGGSAACAKOLLQDLIEQQOOLAKKEIEBELKSERDTLLARIEMRRMQ 269
Db      634 GGGPPALPE-----DLVININSSDEEBEGEEEBEEEEEE-----EEEEEE 682
QY      270 LVKDNKERHKLFGYTE-----EREETELSEKIKLECOPELSETSOTLPPKPPS 321
Db      683 EDFEEDEDEEYFEEEBEEFEFEFEELBEEBEEBEEBEEBEEV---EDLE 739
QY      322 CGRSGGHRKRSFSGTERKTPVKTLAPFSKYKTPKHSPIKEBPCSLSETVCKREL 381
Db      740 FGTA-----GEVEGAPPPPLPALPPSPKVPQPEPEPGLLEV-----784
QY      382 RSQETPEKPRSSVDTPPRLS-----TPQGPSTHPKEKAFSSEIDLPYL 426
Db      785 --EPPTEERGADTAPTLAPALPSOGEVERGESPAAGP---PQELVEEPSXPPTL 839
QY      427 STTEMYLCRWHQPPPLPLRESSPKKEETVARCLMPSSVAGETSVLAVSWRHSVEPL 486
Db      840 LEEETEDGSKVQPPPETPAEE--METETEAALQEKQDDTAAML--DFIDCPDDE 895
QY      487 RDPNPSD 493
Db      896 KPPPTPE 902

RESULT 2
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,014
; FILING DATE: 19-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/574,959
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

Query Match      6.2%; Score 199.5; DB 4; Length 905;
Best Local Similarity 21.9%; Pred. No. 1.3e-05;
Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22;

QY      38 PGAAENHFLPRRK-----LKEGEPPLAASOGSGPAPSGCGKGGLLPAGANG 90
Db      414 PTPAPVPLLRPHRPSGPHRSILRAPCPQWA-----PCPQOAPCPGAG-----PMPGAP 462
QY      91 QQESWG-----SVLPQPP-----PATKQIGGPPAAGAGCS-----126
Db      463 VPSEPTSTTANLGLLSRPSVCPRLPLGPENHRAGSNEDPLAPSGTPPTIPDETFF 522
QY      127 ---PRPKY-----QAVLPLOTGSLVAANE-----PTPMAGDKG 157
Db      523 GGRVPRPAVHYHKEASDVEISLSDSDSVIYVEGLPLPPPPSGATPPPIAPTGP 582
QY      158 GAASPAATASD-----PAGPPPLPLPGPP---LAPTATAGTLAASGKMKSPPLG 209
Db      583 PTASPVPAKKEPEELPAAPGLPPPPPPPPVGVXLLPPQLVPEG-----TP--G 633
QY      210 GGGSGGSAACAKOLLQDLIEQQOOLAKKEIEBELKSERDTLLARIEMRRMQ 269
Db      634 GGGPPALPE-----DLVININSSDEEBEGEEEBEEEEEE-----EEEEEE 682
QY      270 LVKDNKERHKLFGYTE-----EREETELSEKIKLECOPELSETSOTLPPKPPS 321
Db      683 EDFEEDEDEEYFEEEBEEFEFEFEELBEEBEEBEEBEEBEEV---EDLE 739
QY      322 CGRSGGHRKRSFSGTERKTPVKTLAPFSKYKTPKHSPIKEBPCSLSETVCKREL 381
Db      740 FGTA-----GEVEGAPPPPLPALPPSPKVPQPEPEPGLLEV-----784
QY      382 RSQETPEKPRSSVDTPPRLS-----TPQGPSTHPKEKAFSSEIDLPYL 426
Db      785 --EPPTEERGADTAPTLAPALPSOGEVERGESPAAGP---PQELVEEPSXPPTL 839
QY      427 STTEMYLCRWHQPPPLPLRESSPKKEETVARCLMPSSVAGETSVLAVSWRHSVEPL 486
Db      840 LEEETEDGSKVQPPPETPAEE--METETEAALQEKQDDTAAML--DFIDCPDDE 895
QY      487 RDPNPSD 493
Db      896 KPPPTPE 902

RESULT 3
US-08-574-959A-7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:11 ; Search time 28 Seconds
(without alignments)
645.203 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3238

Sequence: 1 MTWRSVAFKAAAPAGGNPE.....RSRCRLTEIKKQTPHRTCRK 614

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199.5	6.2	905	2	US-08-574-959A-9
2	189.5	6.2	905	4	US-09-357-014-9
3	189.5	6.2	1135	2	US-08-574-959A-7
4	199.5	6.2	1135	4	US-09-357-014-7
5	174	5.4	527	4	US-09-370-838-216
6	173	5.3	1384	4	US-08-976-255-11
7	171	5.3	543	4	US-09-535-008-63
8	171	5.3	577	4	US-09-535-008-61
9	171	5.3	1646	4	US-09-535-008-67
10	171	5.3	1647	4	US-09-535-008-75
11	171	5.3	1650	4	US-09-535-008-71
12	171	5.3	1678	4	US-09-535-008-69
13	171	5.3	1678	4	US-09-535-008-65
14	171	5.3	1679	4	US-09-535-008-77
15	171	5.3	1681	4	US-09-535-008-73
16	171	5.3	1682	4	US-09-535-008-73
17	169.5	5.2	3969	4	US-08-061-376-5
18	164	5.1	1618	1	US-07-853-913-4
19	162	5.0	1593	4	US-08-628-829-4
20	160.5	5.0	1185	4	US-09-041-886-23
21	160	4.9	1137	1	US-08-369-043-2
22	158.5	4.9	504	4	US-09-219-849-3
23	158.5	4.9	561	1	US-08-642-255-52
24	158	4.9	960	4	US-09-219-849-5
25	154.5	4.8	1493	4	US-09-423-890-8
26	154	4.8	720	4	US-09-219-849-4
27	154	4.8	777	1	US-08-642-255-53

28	152.5	4.7	234	1	US-08-642-255-51	Sequence 51, Appl
29	152.5	4.7	1064	1	US-08-642-255-62	Sequence 62, Appl
30	151.5	4.7	2414	1	US-08-227-536-2	Sequence 2, Appl
31	151.5	4.7	2414	5	PCT-US95-04682-2	Sequence 2, Appl
32	151	4.7	1060	3	US-08-931-820-3	Sequence 3, Appl
33	151	4.7	1418	3	US-08-963-825-20	Sequence 20, Appl
34	151	4.7	1418	4	US-09-010-999-1	Sequence 1, Appl
35	151	4.7	1418	4	US-09-500-811-20	Sequence 20, Appl
36	151	4.7	1418	4	US-09-570-573-20	Sequence 20, Appl
37	151	4.7	1418	4	US-09-548-608-20	Sequence 20, Appl
38	150.5	4.6	1274	4	US-09-095-443-2	Sequence 2, Appl
39	149.5	4.6	529	1	US-09-247-806-2	Sequence 2, Appl
40	149.5	4.6	595	1	US-08-425-069-4	Sequence 4, Appl
41	149.5	4.6	595	2	US-08-317-844B-4	Sequence 4, Appl
42	149.5	4.6	902	1	US-08-396-479B-6	Sequence 6, Appl
43	149.5	4.6	902	1	US-08-818-823-6	Sequence 6, Appl
44	149.5	4.6	2441	1	US-08-194-468-2	Sequence 2, Appl
45	149.5	4.6	2441	3	US-08-961-739-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-574-959A-9
: Sequence 9, Application US/08574959A
: Patent No. 5962224
:
: GENERAL INFORMATION:
: APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
: APPLICANT: and Jack L. Strominger
: TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
: TITLE OF INVENTION: AND USES THEREFOR
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESS: LAHYE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/574,959A
: FILING DATE: 19-DEC-95
: ATTORNEY/AGENT INFORMATION:
: NAME: Mandragouras, Amy E.
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: DEN-008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 905 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-574-959A-9
:
: Query Match 6.2%; Score 199.5; DB 2; Length 905;
: Best Local Similarity 21.9%; Pred. No. 1.3e-05;
: Matches 120; Conservative 53; Mismatches 225; Indels 149; Gaps 22;
:
: Oy 38 PGAAEAHFLPRHRK-----LKPPGPLSSQGGSPAPSPAGCGGKRGILLPAGAPG 90
: Db 414 PTPAPVLLRPHRSRHSRLRAFCPQMA-----PCQOAPCPSSAG-----PMSAGP 462
: Oy 91 QQESWGG-----SVPLPCPP-----PATKACIGGEPAAAGACS----- 126
```



```
; Sequence 4, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DVL
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-307-019-4
```

```
Query Match          53%; Score 170.5; DB 9; Length 1610;
Best Local Similarity 22.1%; Pred. No. 0.33;
Matches 134; Conservative 51; Mismatches 203; Indels 217; Gaps 30;
```

```
QY 10 AAAAPAGNP--EORLDYERAA-----ALGPEDEPG---AAEAHFLPRHRLKKEBG 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 AGALPGLREPLMEHRYLEEAAREGATILAKAPSFETLRPLPAGSTHLPQSHSLEHD 609
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 57 PPLASQGGSPASPAGCCGCGKGLLPPGAAP-----GQGEESWGSVPLPCPP 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 610 SP-----STPRSSACGAEQRLPSAPSGAPIRDMGHQSGKQLPSTGCH-PGTAQPE 662
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 ATKQAGIGGEPA-----AAGAGCSPPRYQAVLPQTGSL--VAAAKE-----PTW 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 RSPSPWQGPAPFCHPKQGSAPQEGCSFHP--AVAPCPGSPFGSCKEAPLVSSPF 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 ACDKGAASPAATASDP---AGPPPLPLPPLPPLATATAGTLA--ASEGRWKSMSKSP 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 720 LQQPAPAPAP-AKASPPIDSKMGPGDISLPGRPKPGPCSPGSAQASSQVSSILRVSS 778
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 GGGGSGGASSQAACQKILLQLDLLEQQOOLAKKEIEELKSERDTLLARIERMERR 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 QYGTGPGSLDA-----EGWTQEAEDLSDSTPTL----- 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 KQLVKKDNEKERHKLQ-----GYET-----EEREE 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 ---ORPOEQATMRKFSLGGRGYAGVAGYGTFAFGDGAGMLGQPMWARIANAVSOS 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 294 TELSEKILCOPELSETSQLPPKPSGSGRSGKHRRKSPFGSTERKTPYKKLAPFSK 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 EEDDEARAESEEQDEARAESEPLPOVSAR-----PV-----PEVGR 901
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 354 VKTTPKHSPIKEEPCGSLSETVCKREL---RSQEMPEKRSVDTPPRLSTPOKGPST 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 APTRS-SPEPTPWEDIGVS-LVQIRIDSGDAEADTISLDISVD--PAYLNT----- 951
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 410 HPKEKAFSSETEDLPLYSTEMYLGRWHQPPPLPLRESSPKKETVARCIMPSSVAGE 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 ---SDLYDIKYLPEFEMI---FRKYPKSAQP-EPPSPMAEEELAE----- 989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 470 TSVLAVPSMRHSHVEPLRDPNPSDLLENLDSVFSKRAKLELDEK----- 515
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 990 ---PPEPTW-----PWPGL-----GPHAGLEITESESDVDALLAEAAVG 1026
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 516 RKRKM 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1027 RKRKM 1031
```

Search completed: July 14, 2003, 12:00:17
Job time : 58 secs

```
Qy      597 RCRLEIQKKQTPHRTCR 613
      | ::| ::| :
Db      602 EMRSVLRKVGSPRKARR 618
```

```

RESULT 13
US-09-824-574-4
: Sequence 4, Application US/09824574
: Publication No. US20030077800A1
: GENERAL INFORMATION:
: APPLICANT: Rouleau, Natalie
: APPLICANT: Mollanen, Ann-Mariit
: APPLICANT: Palvimo, Jorma J.
: APPLICANT: Jime, Olli A.
: TITLE OF INVENTION: ARIP4 Gene and Protein
: FILE REFERENCE: 2630-109
: CURRENT APPLICATION NUMBER: US/09/824,574
: CURRENT FILING DATE: 2001-04-03
: NUMBER OF SEQ. ID NOS: 7
: SOFTWARE: PatentIn version 3.0
: SEQ. ID NO. 4
: LENGTH: 1647
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-824-574-4

```

Query Match	5.3%;	Score 171;	DB 9;	Length 1647;
Best Local Similarity	19.9%;	Pred. No. 0.32;		
Matches 113;	Conservative 65;	Mismatches 249;	Indels 140;	Gaps 21

```

OY 5 LKPRGRYLSAOGSGPAPRACSGGKGRILL--PAGAQAQOEBSMGSGVPLRCPRA-108
Db 1 MSTRPDRPL-----GGTRPGRSPGRPRSPQAMIGSPGSPGSAHSMG--PSPGPRSAG 53
OY 109 --TKOAGIGEP-----AAAGAGSGPREYUOAV--LPJOTGSLYAAAKERTPW 152
Db 54 HPJRTGQREGYRDQDNHNMHKRMZMNEKMSGSDDERVYQNMGMMSRSGNAGMKPRSPM 113
OY 153 AGCKGAABRATASDPACRPRLPLRGRRPLAPTAGTASLASEGKMSKMSPLSGGGG 212
Db 114 DOHSOQYPSPLGGSEHASSPVPASPSGSP-----QMSGSP--GGAPL 154
OY 213 SGASAOACILQJLLDLDLEOQOQOQOQOAKKELELKSF--KDLIARIEREKRMOL 270
Db 155 DGDAPDA-----LGOQNGPRPPNNOJHOIAQIMAYUKILAKOPRLDHL 202
OY 271 VKKDNEKERHKLFGQYETEEREETELSEKIKLEQPELSETSOTLPRKFGSGRSGK-- 348
Db 203 A-----VQGRPR-----MRGQOQMPLPLRPSPVSATGPRGPRG 235
OY 329 -----HKRSPFGSTERTKTPVKKILAPESVUKTRKNSPIKEEPCGSLSET 375
Db 236 PGDPRGPRAPRYSPRHMGGRPMRPGSPGVPGRMGQ---PBGPRKPMGPRMANA 292
OY 376 VCKRELQSETRPKRSSVDTPRLSTQKQSGTHPKKAKSSLELDDPLYSTTEMLCR 435
Db 293 AAPTSTQKILIPRQGRPSRPARVPAPASVMPROTQSPGQAPAPMPL----- 345
OY 436 WHOPRPSPLP-----LRESSPKKEETVARCL-----MPSVAGETSVLAVPSW 478
Db 346 -HQKQRIPIPIQKPRGLDREVELLOREYRLQARINHRIOELENLPGSLAGDLRKAI-- 401
OY 479 RDHSVERLDRPNBSOLLNLNDVSFSKRNHAKLE--LDEKRRKMDIORIREORILLQOL 536
Db 402 --TELALRYLNEQROLRO--EYUVCVMRDTALETLNAAKYSKROSLREARITTEK 458
OY 537 RM---YKKKGIOSESEPEVVSFFERDD 560
Db 459 QOKIEQERRKROHDEYLSIJLQNAKD 485

```

```

Sequence 1, Application US/10307019
Publication No. US2003010853A1
GENERAL INFORMATION:
APPLICANT: Zeng, Weilin
APPLICANT: Stanton, Lawrence
APPLICANT: SCIOS, INC.
TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
FILE REFERENCE: SCIOS.021bV1
CURRENT APPLICATION NUMBER: US/10/307,019
CURRENT FILING DATE: 2002-11-26
PRIOR APPLICATION NUMBER: 09/548,473
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/129,552
PRIOR FILING DATE: 1999-04-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1351
TYPE: PRT
ORGANISM: Homo sapiens
US-10-307-019-1

```

Query Match	5.3%;	Score 170.5;	DB 9;	Length 1351;
Best Local Similarity	22.1%;	Pred. No. 0.27;		
Matches 134;	Conservative 51;	Mismatches 203;	Indels 217;	Gaps 30;

```

OY      10  AAAAPRGVN- -BORTUDYRAA- -ALGPRDEGG- -AAEAFLPRNHKLTKEPG 56
Db      291  AGALPGIREIMENHRILEEBAAREQATILKARFETALRLPSGTHLARGHSHLEHD 350
OY      57  PPLASSOGSPABSPAGCGGKRGILLPAGAA- - -GOEBSGGSVPLRCPRP 107
Db      351  SP- - - - -STPPRSSACGEAORILPSAPSGGARDIMHMGOSQOLPSTGGH- -PGTAQPE 403
OY      108  ATKQAGIGGEP- - - - -AAGACSPRPYUAVLITQSG- -VAALKE- - -PTPW 152
Db      404  RPSRDSBWCQPAFCFCHPKQGSARQEGCSPH- - -AVACSPRGSPRPGSCKRALVPSPF 460
OY      153  AGCKGAASPAATASDP- - -AGPPPLPRPPPLATPATAGTLA-ASEGKMSKMSPL 207
Db      461  LGOPQAPPA- -AKASPPIDSKMGPDISLPGKPKGPCSSPGSASQSSSVSSSLRVGSS 519
OY      208  GGGGGGASSQAACQILLLDLDLIEQOQOOLAKKEIEELSERDITLIERMERR 267
Db      520  QVTEBPGPSLDA- - - - -ECWTEARDLSOSPRTL- - - - -548
OY      268  MOLVKDNEKERIKLQ- - - - -CYE- - - - -EBREE 293
Db      549  - - - - -QROEATMRKFTSLGRCGVAGVAGTGFARFAGDAGMLGQPMARIAMAVQSE- 604
OY      294  TELSEKIKLECOPELSETQTLPRPKFSGRSGKHKRKSPPGSTERKTPYVKLAPFEK 353
Db      605  EEEOEEARASQSEEOQEARAESPLRYQSAR- - - - -PV- - - -PEVGR 642
OY      354  VKTKTKHSPKIKEPGSGSETVCKREL- - -RSQETPEKRSADVPPRLSTPQKQST 409
Db      643  APTRS- -SPETTPMEDIGQVS- -LVQIRDSGAEMADTISLDISEVD- -PAYINL- - - -692
OY      410  HPKEKAFSSEIEDLPYLSTTEMYLCRHQOPPPSPLPRESSPKKEETVYARCIMPSSVAGE 469
Db      693  - - - - -SDLYIKILPPEFMI- - -FRKVPKSAQ- -EPSPMAEEELAE- - -730
OY      470  TSVLAVPSMWDHVEPLRDPNPSDLELNDQSVFKRAKLELEK- - - - -515
Db      731  - - -FPRETW- - - - -PWPGEU- - - - -GPAAGLEITEESDQVALLAEAAVG 767
OY      516  RKRKM 520
Db      768  RKRKM 772

```

RESULT 14
US-10-307-019-1

RESULT 15
US-10-307-019-4

```
Qy      481 HSEVPLRDPN--PSDILEN 497
          | | | : | : ||
Db      512 ---PPARAPDARPAQPVEN 527
```

RESULT 11
US-09-919-039-278

```

? Sequence 278. Application US/0919039
? Publication No. US20030108871A1
? GENERAL INFORMATION:
? APPLICANT: Kaser, Matthew R.
? TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
? FILE REFERENCE: PA-0035 US
? CURRENT APPLICATION NUMBER: US/09/919,039
? CURRENT FILING DATE: 2002-09-09
? PRIOR APPLICATION NUMBER: 60/222,113
? PRIOR FILING DATE: 2000-07-28
? NUMBER OF SEQ ID NOS: 401
? SOFTWARE: PERL Program
? SEQ ID NO 278
? LENGTH: 707
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
? US-09-919-039-278

```

Query Match	5.38	Score 172.5	DB 9	Length 707
Best Local Similarity	22.38	Pred. No. 0.1		
Matches 137; Conservative	58	Mismatches	200	Indels 219; Gaps 32

QY	49	HRKJKEPPPLASS-----QSGSPSPAGCGKRGILLP-----GAPO	91
Db	94	HQOQOQPPRPDDSKRPVVAQGBRP-----GVGASPRPSSSAPRPITPSGAPG-	144
QY	92	QEESMGSVLPRLCPRPATKOAGIGGEPAAAGCSPPRYQAVLPLOTGSLVAAAKEPT-	150
Db	145	-----SGGPPPTPRPAVTSAPRGAP-----PTPSSGV-----PTT	177
QY	151	-PMAGDKGASAPATADPAGRPRLPRP-----PLAPNATGTLAASGRKMSMK	204
Db	178	PRQAG--GPPPPRAVPRGPBPCKGPGGKGKGMKPGKPRGGGPGJSTG--GHPK	232
QY	205	SPLGGGSGSASSQAACLQIILLDLBLEQOQO-----LOAKEKEIETELSERDT	256
Db	233	PRHGGGPRGRGQ-----HPRYHQHNGGPRPGGSGGSEK-----	271
QY	257	LIAIERNERNMOLYKKNK-----ERHKLFG--YETEERETELSEK-----	299
Db	272	-ISDSEGFKAMLSILRRPGKTYTORCLFVGNLPADITEDFFKRLFAUYGGEVFINK	330
QY	300	-----IKLEQPELSETSOFLPRK-----FSGRSGSKGKKRSP-----	334
Db	331	GKGFGFKLESRALAEIKAEILDTPMRGRDLRVFATHAALSVRNLSPIVSNELLEEA	390
QY	335	-----FGSTER-----KTPVKLIAPEFSK-----VKTKTPKHSPIKEE	367
Db	391	FSQGPRIERAVIYDDRGSTGKGIVFEASPKAPAAKAPERCSEGYLLTTTTR--PYIYE	448
QY	368	PCGSLSETVCKRELRSQETP--EKPRSSVDPPRLSTPOKGPSTHPKREKAFSSIEDLPY	425
Db	449	PLEQDDEDDGJPEKLAQKNPMYOKER--ETPPREA--OHG-----TFEYE-----	489
QY	426	LSTTEMYLCRHHORPPSLPLRESSPKKEEYARCLMSSVAGETSVLAVPSMRDHSVER	485
Db	490	-----YSQWRKS-----LDEMEKQOREYKMKDAKDLSEEM--EDAYHEHOANL	534
QY	486	LADENPSDL-----LENLDSDVFSKR--HAKLEJDEKRRKMDIORIRORI-----LO	532
Db	535	LR-----ODIMRQOEELRRMEELHNOEMOKRKEMOLROEBERRRRREEMNITROREMEOR	590
QY	533	RLQDRMYKKKGIOE	546

Db 591 RQREESYSRMGYMD 604

RESULT 12
US-10-171

```

: Sequence 1 Application US/10171384
: Publication No. US20030031680A1
: GENERAL INFORMATION:
: APPLICANT: The Brigham & Women's Hospital, Inc.
: APPLICANT: Belter, David
: APPLICANT: Herion, Bruce
: TITLE OF INVENTION: p53 Binding Protein-Related Protein in Cardiomyopathy
: FILE REFERENCE: 81994/275368
: CURRENT APPLICATION NUMBER: US/10/171,384
: CURRENT FILING DATE: 2002-06-14
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 824
: TYPE: PRT
: ORGANISM: mouse
: US-10-171-384-1

```

Query Match	5.3%	Score 172;	DB 9;	Length 824;
Best Local Similarity	21.9%	Pred. NO. 0.13;		
Matches 148;	Conservative 56;	Mismatches 230;	Gaps 30	

[illegible]

RESULT 9
US-10-144-649A-216
; Sequence 216, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Tonglong
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: THE THERAPY AND METHODS FOR
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-144-649A-216

Query Match 5.4%; Score 174; DB 9; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA-----ALGPPDEPGAEE-AHFLLRHRKLT-----BPGPLIASOGGSPAPSP 71
DB 8 EKAENGALGSPREKEVLENGELTPPRREKALENGELSPAGKVLVNGSLTPPKSE 67
QY 72 ACCGGKRGILLPAGA-----APGOQE--BSWGSVPL--PCP-----PP 107
DB 68 DKVSENG-GLRFRNTERPETGPMRAPGWEKTPSMGPAITIGEPAPETSLERAPAPS 126
QY 108 ATQAGTIGEPAAAGACSPRKYQAVLP-----IQTGSLVAAKEPPTWAG--DKG-G 158
DB 127 AVVSSRNGETAGPLG--PAPKNGTLEGTERRAETGAPARA-----PGAGRLDIGSG 179
QY 159 AASPATYASDPAG-----PPPLPLGPP----- 181
DB 180 GRAPVGTGTAPGGGPGSSVDKAGWYDNTRPQPPPLPPPPPAQRRLEPAPRAPPEV 239
QY 182 -----PLAPYATAGTLAASEGRW--KSMRKSPLGGGSGGASSQAACLKQILLQDLI 233
DB 240 APEGEGAPDSRAGGDTALSGDDGPPKPERKGP-----EMRPLFLDIG 282
QY 234 EQQOQOOLQAKKEIEELKSRDILLARIEMERRMOLVKKDNKEKHKLFQGIET----- 288
DB 283 PPOGENSEQIKAR-----LSRLSLALPPLTLTPPGGPRRPWEGADAGAGG 330
QY 289 -----DERETELSEKIKLECOPELSETQTLPPKPFSCGSGKGRKSPF- 335
DB 331 EAGGAGAPGAEDGEDDEDE-----EDEEAAPGA--AAGPRPGARAPAPV 379
QY 336 -----GSTERKTPVKTLAEPFSVKTKTPKHSPIKEEPCGSLSETVCKRELRS----- 383
DB 380 VVVSADADAPARLRL-----LKSPPGADEPD-----SELERKKWVSFHGDVTV 426
QY 384 -----QETPEK-----PRSSVD-----TPRLSTPOKG--PSIHKPKAASSEL----- 420
DB 427 YLFDETPNELSVQAPPEGDTDPSTPPAPPTPHATPGDGPFS--NDSGGGSFEWA 483
QY 421 EDLPYLSTEMYLCRWHOPPPSPPLPRESSPKKEETVARCLMPSSVAGETSVLAVPSWRD 480
DB 484 EDFPPL-----PPPGP-PL-----CFSRFSV--SPALETG----- 511
QY 481 HSYVEPLRDPN--PSDILEN 497
DB 512 ---PPARAPDARAPGIVEN 527

RESULT 10

US-09-738-973-216
; Sequence 216, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indritas, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-216

Query Match 5.4%; Score 174; DB 10; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 186; Indels 246; Gaps 36;

QY 26 ERAA-----ALGPPDEPGAEE-AHFLLRHRKLT-----BPGPLIASOGGSPAPSP 71
DB 8 EKAENGALGSPREKEVLENGELTPPRREKALENGELSPAGKVLVNGSLTPPKSE 67
QY 72 ACCGGKRGILLPAGA-----APGOQE--BSWGSVPL--PCP-----PP 107
DB 68 DKVSENG-GLRFRNTERPETGPMRAPGWEKTPSMGPAITIGEPAPETSLERAPAPS 126
QY 108 ATQAGTIGEPAAAGACSPRKYQAVLP-----IQTGSLVAAKEPPTWAG--DKG-G 158
DB 127 AVVSSRNGETAGPLG--PAPKNGTLEGTERRAETGAPARA-----PGAGRLDIGSG 179
QY 159 AASPATYASDPAG-----PPPLPLGPP----- 181
DB 180 GRAPVGTGTAPGGGPGSSVDKAGWYDNTRPQPPPLPPPPPAQRRLEPAPRAPPEV 239
QY 182 -----PLAPYATAGTLAASEGRW--KSMRKSPLGGGSGGASSQAACLKQILLQDLI 233
DB 240 APEGEGAPDSRAGGDTALSGDDGPPKPERKGP-----EMRPLFLDIG 282
QY 234 EQQOQOOLQAKKEIEELKSRDILLARIEMERRMOLVKKDNKEKHKLFQGIET----- 288
DB 283 PPOGENSEQIKAR-----LSRLSLALPPLTLTPPGGPRRPWEGADAGAGG 330
QY 289 -----DERETELSEKIKLECOPELSETQTLPPKPFSCGSGKGRKSPF- 335
DB 331 EAGGAGAPGAEDGEDDEDE-----EDEEAAPGA--AAGPRPGARAPAPV 379
QY 336 -----GSTERKTPVKTLAEPFSVKTKTPKHSPIKEEPCGSLSETVCKRELRS----- 383
DB 380 VVVSADADAPARLRL-----LKSPPGADEPD-----SELERKKWVSFHGDVTV 426
QY 384 -----QETPEK-----PRSSVD-----TPRLSTPOKG--PSIHKPKAASSEL----- 420
DB 427 YLFDETPNELSVQAPPEGDTDPSTPPAPPTPHATPGDGPFS--NDSGGGSFEWA 483
QY 421 EDLPYLSTEMYLCRWHOPPPSPPLPRESSPKKEETVARCLMPSSVAGETSVLAVPSWRD 480
DB 484 EDFPPL-----PPPGP-PL-----CFSRFSV--SPALETG----- 511

Query 367 -----BPCSLSETVCKRELRSQETPE-KPRSSVDTRP----- 398
Db 991 KPVPPVPPPHLOPBEEDVSO-----QSGSGPRGKRSRV--PRAKEAKAPFAAF 1041
Query 399 -----RLSTPQKG---PSTHPREKAFSSIEDLPYLSTTEMYLCRMHOPPSPPLP 446
Db 1042 PTEGOSYRLS-PHAGHRLPSHPREVIKTSTRAD-PLFSYV-----PGRHPLPL 1088

RESULT 7
US-09-862-027-80
; Sequence 80, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345,473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 1265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-027-80

Query Match 5.4%; Score 174.5; DB 10; Length 1265;
Best Local Similarity 23.6%; Pred. No. 0.15;
Matches 134; Conservative 51; Mismatches 201; Indels 183; Gaps 28;

Query 4 RSANFKAAAPAGNPPQRLDYERAAALGPEDEPGAENFLPRHKKLKPGLA--- 60
Db 759 RSSLFRTKQASLLHTSRLSLNRSLSGSGSPGS-----PTHSHSLSPRSPTQYR 812
Query 61 -----SSGGSPA-----PSPAGCGKRGILLPAGAPGOOE-----SWG 98
Db 813 VTPPAVHVGNSQSSPSSSVSPFAGSHTPPSSL--HGLAPKLDQRRSPRRKSAG 870
Query 99 SVPL-----PCPPATKQAGIGGEPAAGACSPRPKYQAVLPIDTGSILVAAKPTPW 152
Db 871 SIPLSLAHTPPSPPT-----ASPO-----RSPSL 897
Query 153 AGDGGAAPATASDAGP-PLPLPGPPPLAP-----TATAGTLASEGRKMSR 203
Db 898 SGHVAQAFPTKLLHSPPLGRSLRPSKAEPRSPPLKRVQSAEKLAAALASEKKLATSR 957
Query 204 KSPAGGGGSGASSQAACLQILLQDLLEOQ--OOQLOAKEKEIEELKSERDTL----- 257
Db 958 KH-----SIDLPHSELKELPREVSPLEVVAGARVLSGKA 994
Query 258 -----LARIEMRMOLVKKDNEKERHKLPGQYETEE-RETELSEKI 300
Db 995 LPGAQVLOPAPSRALGTLRDRARRRSLOKQEAIREVDS--SEDDTEEGEENSGAOEL 1052
Query 301 KLECQPELSESOFLPPKPFSCGSGKGRKSPGSTERKTPVKTLAPESXYKTKTPK 360
Db 1053 SLAPHEV---SOSVAPK--GAGESGE---EDFPFS---RGP-RSLGPMVPSLTLGTL 1099
Query 361 HSPKIEEPCGSLSTVCKRELRSQETPEKPRSSVDTPPRLSTPQKGFTH--PKE---KAF 416
Db 1100 GPPMESPSPG-----HRRIGSPQATEEAASSSSAGNCG--QSGADTLPPEGCMAQ 1151
Query 417 SSEEIDLPYLSTTEMYLCRMHOPPSPPLPRESSPKKEETVYARCLMPSSVAGETSVLAVP 476
Db 1152 HLHTQALTAIS-----PSTSGTLPTS-----SCSPSSTSGK---LSMW 1187
Query 477 SMRD-----HSVEPLRDPNPSDLEND 500
Db 1188 SWSKSLIEGPDRASSPRKATYAGGLANLQD 1216

Query 8
US-09-854-133-216
; Sequence 216, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-216

Query Match 5.4%; Score 174; DB 9; Length 527;
Best Local Similarity 23.1%; Pred. No. 0.061;
Matches 143; Conservative 44; Mismatches 166; Indels 246; Gaps 36;

Query 26 ERAA---ALGPEDEPGAEE-AHFLPRHKKIK-----EPGPASSGGSPAPSP 71
Db 8 EKAENALGSPEREEREKYLENGELTPPRREKALENGELRSPEAGEKYLVLNGGLTPKSE 67
Query 72 AGCGKRGILLPAGA-----APGOE---ESMGSGVPL--PCP-----PP 107
Db 68 DKVSENG-GKREPNTERTPEPTGPWRAPGPEWKEPESMGPAVPTGEPAPETSLERAPAP 126
Query 108 ATKQAGIGGEPAAGACSPRPKYQAVLP-----IOTGSILVAAKBPPTWAG--DKG-G 158
Db 127 AVSSRRNGGERAPPLG--PAKNGTLEPGERARPELGGAPRA-----PGAGLIDLGSG 179
Query 159 AASPAATASDPAG-----PPPLPGPP----- 181
Db 180 GRAVGTGTAPAGGPGSGVDKAKAGVNDTRPQPPPLPPEPEAOPRLERAPRARREV 239
Query 182 -----PLAPVTAGTLAASEGRW---KSMKSPGSGGSGASSQAACLQILLQDLI 233
Db 240 APGEPPAPDSRAGDRLASDGDPPKPERKGP-----EMRFLFDLG 282
Query 234 EQQOQLOAKEKEIEELKSERDTLLARIEMRMOLVKKDNEKERHKLFGQYET----- 288
Db 283 PPGNSSEQIKAR-----LSRLSLALPLTLTPPGPGRPRRPMEGADAGAAG 330
Query 289 -----ERRETELSEKIKLECQPLSETSQTLPPKPPSCGSGSGKHKRKPFP- 335
Db 331 EAGGAGAPGAEEGEDEDEDE-----EDEEAAAGA--AAGPRGGRARAAPVP 379
Query 336 -----GTERKTPVKTLAPESKVTPTPKPSPIKEPCGSLSTVCKRELRS----- 383
Db 380 VVVSADADAARPLRGL-----LKSFRADRPED-----SELEKRRKRVSHGQYV 426
Query 384 -----QETPER-----PRSSVD-----TPRLSTPQKG-PSTHPREKAFSSEI--- 420
Db 427 YLFOETPPTNLSVQAPPEGDTDPSTPPAPTPPHAPATPGGFPs---NDSGFGGSPFMA 483
Query 421 EDLPYLSTTEMYLCRMHOPPSPPLPRESSPKKEETVYARCLMPSSVAGETSVLAVPSRD 480
Db 484 EDPPIL-----PPGP-PL-----CPSRFSV---SPALETGP--- 511
Query 481 HSVEPLRDPN--PSDLEND 497
Db 512 ---PPARAPDARPAQVEN 527

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 36182
LENGTH: 617
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005529.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUATE 2.00e-33
OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUATE 3.00e-22
US-09-864-761-36182

Query Match 6.1%; Score 199; DB 10; Length 617;
Best Local Similarity 23.6%; Pred. No. 0.0034;
Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

QY 90 GQOESMGSVPLPCPPPRKQAGIGEPAAAGCSPRYQAVLP-OTGSLVAAKE 148
D 86 GEEEAAGGGEETKSP-----AEFAAS-----PEKEAKSPKEAKSPAEKKS 129
QY 149 PTPMAGDKGGAASPAAT-----ASDPA-----GPPPLPLPPLPPLATAGTAAEGR 198
D 130 P-----EKEAKSPAEEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 184
QY 199 WKSRRKSPGCGGSGGASQAACLQILLQDLIEQOQOQLOAKKEIEELKSERDTLL 258
D 185 AKEAKSPAEEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 226
QY 259 ARIRMRRLQVKKDKNEKHKLFQGYETEEEREETELSEKIKLECPLESTQTLPRK 318
D 227 AEVKSPEK-----AKSPTEKAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSP 281
QY 319 PFCGSRGKGHRRKSPGSTERKTPVKKLAPEFSKVTKTTPKHSPIKEE-PCGSLSETVC 377
D 282 K---AEAKSPKAKSPV-KEEAKSPKAKSPVKEEAKSPKAKSPVKEEAKSPV 337
QY 378 KRELRSQE---TPKPRSSVDTPRLSTPOKGPSTHP---KEKAFSSEIEDLPYSTTE 430
D 338 KEKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSPKAKSP 397
QY 431 MYLCRMHQPPSPPLRKE---SSPKKE---TVARCLMPSSVAGETSVLAIVSW---BDHS 482
D 398 SPLAADAAPKEKELPKKEEVKSPVKEEKEKPOEVKVEPKKAEKKAAPRPTKEEKDOK 457
QY 483 VE---PLRDPNPSDLLENLDSVFESKRHALELDEKRRKRWIDORI---REQRILOLRM 538
D 458 KEKAPKKEAPKPKVEEKEKKEAPKVEKESKVE---AKKEAEKDKKVPPEKAPAKVE-- 513
QY 539 YKKKGIOSESEPTVSFPEPDVDES 563
D 514 -KEDAKPKREKTEVAK--KEPDADA 535

RESULT 4
US-10-156-761-11739
Sequence 11739, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11739
LENGTH: 1479
TYPE: prt
ORGANISM: Streptomyces avermitilis
US-10-156-761-11739

Query Match 6.0%; Score 194; DB 9; Length 1479;
Best Local Similarity 20.7%; Pred. No. 0.017;
Matches 145; Conservative 59; Mismatches 203; Indels 292; Gaps 31;

QY 9 KAAAPAGCNEDQRLDYERAAALG-----GEDEPGAERHF---LPRHKKLEKPG-- 57
D 889 RPAATPAOGQPT-----PAQAAPOQVPVGGQGVPGGADAPAGMPVPGTGPQPGQG 943
QY 58 -----PLASSOG-----GSPAPSPAGCGGKGKGLLP-AGAAPGQGESWGSV 100
D 944 AIPHOVAPGVAAAGQVPAQAPQVPVSGQAAGQALAPAPAPAPQAPQVPV- 1001
QY 101 PLPCPPPRKQAGIGEPAAAGCSPRYQAVLP-OTGSLVAAKEEPTPW---AGDKG 157
D 1002 PAQVPPAA--DGSVSGVPSGADPALGTPRAQAQAP-----AAAPQPVMPGAAHATS 1052
QY 158 GAASP-AATASDPAGPPLPLPGP-PLAPATAGTLAA-----SEGRWK 200
D 1053 GAGTPLVAGQPPAPAPQPTAPAPGTPPLPERRAOPRAAOPLPAAEAAPIDPNSTGHA 1112
QY 201 SMRKSPLGGGSGGASQAACLQILLQDLIEQOQOQLOAKKEIEELKSERDTLLAR 260
D 1113 SVRTL---GGVPPARQAAA-----QNKPGAQAOPR----- 1140
QY 261 IERMERRMQLVKKDKNEKHKLFQGYETEEEREETELSEKIKLECPLESTQTLPRKPF 320
D 1141 ----- 1151
QY 321 SCGRSGKHRRKSPGSTERKTPVKKLAPEFSKVTKTTPKHSPIKEEPCGSLSETVCKRE 380
D 1152 QTNCGRRRKLGTG-----PE-----PAERP-----ENARPQ 1180
QY 381 LRSQETPE-KPRSSVDTPPR---LSTPOKGPSTHPKE---KAFSSEIEDLPYSTTEYTL 433
D 1181 PQAGOSPOARPAAGTIDGAGRYAIGAPDKNADEGPPLDGGGAVETADQHPRPVDEL 1240
QY 434 CRWHQPPSPPLPREESSPK-----KEE----- 455
D 1241 -----PPEPL---DNBRRLVWPAPDVTTQALSDRGVRYIVHSREEDVDAQIAEPA 1290
QY 456 -----TVARCLMPSSVAGETSVLAIVSWRDSVEPLRDPNPSDLLENL--DD 500
D 1291 ALFVDPPLGPTITRTALQSLQAAVAAEVPLVTAGLQATREAAVYAGDAVPAVLAKALARD 1350
QY 501 S-----VFSKRHALEL-----DE----- 514
D 1351 SEQHPPRYLLIEEHAETSLALTSTLERRGQVAPAAASDEDAVTLAGQMRPNLVMDLMQV 1410
QY 515 KRRKRWIDQRIEQRILQRLQRLMYKKKGIOSESEPTVS 553
D 1411 RRRRAGIVDWLRANGQLNRPPLVYTTAAVDAELPRLAS 1449

RESULT 5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 14, 2003, 11:50:46 : Search time 56 Seconds
(Without alignments)
1276.460 Million cell updates/sec

Title: US-10-054-935-2
3338
Perfect score: 1 MTRMSAVFKAAPAGGNPE.....RSRCRLFIQKQTPHRTCRK 614
Sequence:

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1446	44.7	281	10 US-09-925-302-655	Sequence 655, App
2	206	6.4	1023	9 US-09-893-519A-14	Sequence 14, App1
3	199	6.1	1479	10 US-09-864-761-36182	Sequence 36182, A
4	194	6.0	2701	9 US-10-156-761-11739	Sequence 11739, A
5	187.5	5.8	2701	9 US-10-171-311-83	Sequence 83, App1
6	176.5	5.3	2462	9 US-09-819-104A-5	Sequence 80, App1
7	174.5	5.4	1265	10 US-09-862-027-80	Sequence 216, App
8	174	5.4	527	9 US-09-854-133-216	Sequence 216, App
9	174	5.4	527	9 US-10-144-649A-216	Sequence 216, App
10	174	5.4	527	10 US-09-738-973-216	Sequence 216, App
11	172.5	5.3	707	9 US-09-919-039-278	Sequence 278, App
12	172	5.3	824	9 US-10-171-384-1	Sequence 1, App1
13	171	5.3	1647	9 US-09-824-574-4	Sequence 4, App1
14	170.5	5.3	1351	9 US-10-307-019-1	Sequence 1, App1
15	170.5	5.3	1610	9 US-10-307-019-4	Sequence 4, App1
16	170.5	5.3	1665	10 US-09-858-664A-2	Sequence 2, App1
17	170.5	5.3	2596	9 US-10-307-019-6	Sequence 6, App1
18	170.5	5.3	2630	9 US-10-077-130-2	Sequence 2, App1
19	170.5	5.3	7968	9 US-10-077-130-5	Sequence 5, App1

20	168.5	5.2	1000	9 US-10-128-714-3305	Sequence 3305, Ap
21	165.5	5.1	802	10 US-09-823-240-2	Sequence 2, App1
22	165	5.1	1555	9 US-10-128-714-3298	Sequence 3298, Ap
23	165	5.1	1832	9 US-10-128-714-8298	Sequence 8298, Ap
24	164.5	5.1	758	10 US-09-904-987-5	Sequence 5, App1
25	164.5	5.1	3256	10 US-09-919-039-21	Sequence 21, App1
26	164.5	5.1	3256	10 US-09-919-172-98	Sequence 98, App1
27	164	5.1	1001	9 US-10-128-714-3240	Sequence 3240, Ap
28	164	5.1	1618	9 US-09-963-875-1	Sequence 1, App1
29	164	5.1	1618	9 US-10-136-891-2	Sequence 2, App1
30	164	5.1	1618	9 US-10-120-687-1	Sequence 1, App1
31	163.5	5.0	810	9 US-10-156-761-10081	Sequence 10081, A
32	162.5	5.0	538	9 US-09-976-740-43	Sequence 43, App1
33	162.5	5.0	538	12 US-10-023-529-43	Sequence 43, App1
34	162.5	5.0	538	12 US-10-023-523-43	Sequence 43, App1
35	162.5	5.0	1259	9 US-10-260-715-8	Sequence 8, App1
36	160.5	5.0	775	9 US-09-738-626-3773	Sequence 3773, Ap
37	159.5	4.9	1312	9 US-10-029-115-2	Sequence 2, App1
38	158.5	4.9	629	10 US-09-833-790-429	Sequence 429, App
39	158.5	4.9	1295	9 US-09-789-390-30	Sequence 30, App1
40	158.5	4.9	1295	9 US-09-789-390-32	Sequence 32, App1
41	158.5	4.9	1295	9 US-09-789-390-34	Sequence 34, App1
42	158.5	4.9	1295	9 US-09-789-390-37	Sequence 37, App1
43	158.5	4.9	1295	9 US-09-789-390-39	Sequence 39, App1
44	158	4.9	824	9 US-09-909-567B-53	Sequence 53, App1
45	158	4.9	912	9 US-09-291-417-26	Sequence 26, App1

ALIGNMENTS

RESULT 1
US-09-925-302-655
Sequence 655, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 655
LENGTH: 281
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-302-655

Query Match 44.7%; Score 1446; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 4.8e-70;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 PPLAPRTAGTTLAASGRKMSKRSPLGGGGGSGASSQAACLOILLDLLEQQOOL 240
DB 1 PPLAPRTAGTTLAASGRKMSKRSPLGGGGGSGASSQAACLOILLDLLEQQOOL 60
QY 241 QAKKEKEELKSERDITLARIEMRMOLVKKDNEERKTLQGYETEERETSEKI 300
DB 61 QAKKEKEELKSERDITLARIEMRMOLVKKDNEERKTLQGYETEERETSEKI 120
QY 301 KLEQPELSETSQLPPKPSGSGKGRKSPFGSTERTKTPVKKLAPFSVKTKTPK 360
DB 121 KLEQPELSETSQLPPKPSGSGKGRKSPFGSTERTKTPVKKLAPFSVKTKTPK 180
QY 361 HSPKPEPCGSLSTVKKRELRSQETPEKRRSSVDTPPRLSTQKGPSTHPKKAASSEI 420
DB 181 HSPKPEPCGSLSTVKKRELRSQETPEKRRSSVDTPPRLSTQKGPSTHPKKAASSEI 240

Tue Jul. 15 10:16:07 2003

us-10-054-935-2.rni

Page 21

Db 441 GCACCACTCCGGCAGCACCAGACCACTCCGGGCGCACTCC-----394
 Oy 458 ALAArgysLeuMetProSerValAlaIgluIrnHisValLeuAlaValProSer 477
 Db 393 -----ATTCCACGGTCATCATCCAGTCCTCATCATCATCGTGCCTCA 343
 Oy 478 TRPArgAspHisSerValGluProLeuArgAspProAspPro 491
 Db 342 TCATAGTCATATTCACATCTCCACCGCGCGCGCTCCG 301

Search completed: July 14, 2003, 23:49:40
Job time : 4383 secs


```

: REFERENCE/DOCKNET NUMBER: F-5610
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (212)966-2340
:
: TELEFAX: (212)953-7733
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 2214
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: ORIGINAL SOURCE:
:
: ORGANISM: Pinctada fucata
:
: CELL TYPE: mantle epithelial cell
:
: US-08-864-038A-1

```

Alignment Scores:

Pred. No.:	0.0102	Length:	2218
Score:	211.50	Matches:	128
Percent Similarity:	34.83%	Conservative:	58
Best Local Similarity:	23.97%	Mismatches:	235
Query Match:	6.53%	Indels:	116
DB:	3	Gaps:	17

US-10-054-935-2 (1-614) X US-08-864-038A-1 (1-2214)

[illegible][illegible]

Db	1866	CTCAGACCCGAGGAGAGAGAGGAACTTCCAGGCGGACCTGGCGGACCTGACTTGTGA	1925
OY	239	nlEglnAlLysgluLysgluLjleLlunLyeuLysSerGluAArgspThrLeuAl	259
Db	1926	GATCGAATATACAGCAGAGCTGTATCCAGACAGCTGGAGAAACACGCGCGCTTGCAGAC	1985
OY	259	anArgLleAArgMetClnAArgArgGlnLeuValLys-----LysAspAnGl	276
Db	1986	GCTCAGACACAGATATAGAGAAAGAGCTGATCTGCTGCAGAAACAAGATCCAGACACACA	2045
OY	276	uLysGlnArgHnLysLeuPheGlnLjYTrcIuThrGluLudArgGluIuThrGluLe	296
Db	2046	CGTGCAGCGCGACCGCTGTCTGCAGAACTCAGACCCATGGAGCTCTATATCAGAGAA	2105
OY	296	uSerGluLysIleLysLeuGluLysGlnProGluLysuSerGluThrHisSerGlnThrLeuP	316
Db	2106	GGCCACCAAGATCAAGCAGACTATGAGAAAGAGCTGGGGAGATGAACGGGACCTGCA	2165
OY	316	oProLysProPheSerCysGlyArgSerGlyLysGlyHnLysArgLysSerPropheG	336
Db	2166	GAAGCTGACGGCGCCCAAAAGACACAGCCCGCTGTTAAAGAACAGTCCGGCTAC--	2223
OY	336	ySerThrgluArgLysThrProValLysLeuAlaProGluPheSerLysVal---Ly	355
Db	2224	-----GAGAGGAG-----CTGAACAACCTACAGCCGAGGTGGCTAGATGAAGAA	2270
OY	355	srThrLysThrProLysHnLysSerProLysGluGlu---ProCysGlySerLeuSerG	374
Db	2271	GGCCACAGTGGCCCTGATGAAGCAGATGCTGAGAGCAACAGGGCGGGCTATGTGA	2330
OY	374	uThrValCysLysArgGluLeu-----	381
Db	2331	GAACCAAGAGAACCCGGAGATCGCACAGCTCAGAAGACAGCGGACAGGAGTTTCA	2390
OY	382	-----ArgSerGlnGlu-----	387
Db	2391	GATCCGAGCTGTGAGTCCAGAGCGCAGACAGAAATGCTCTGAGGAGAAACCCA	2450
OY	387	oGlu-----	396
Db	2451	GGAGGTTTCTGCATGAGGCGCCTGGCCAAAGCCCATGTGACGGGGTGGAGGCGTGC	2510
OY	397	-----ProProArgLeu-----	403
Db	2511	AGGACATAAGCCACCCATGCTGGAGCTGTGGGCTGAGGTGTGCGCCAGACACTACCTATC	2570
OY	404	-----GlnLysGlyProSerThnHisProLysGluLysAlaPheSerSerGluLe	420
Db	2571	TGAGGCTGAATCAGGGGCGCGCTGTCTGTCCACATCTGCGCC-----ACTGGAAACG	2624
OY	421	gluAspLeuProLysLeuSerThnThnGluMetLysLeuCysArgTrpHisGlnProPro	440
Db	2625	CAAAATCAACACACTCTTGGGGGACCATCTGCGCCACTGTCAATATGGCACCCGCTC	2684
OY	441	ProSerProLeuProLeuArgLysuSerSerProLysLysGluGluThrValAlaArgCys	460
Db	2685	CCGAAGAAGATCCAGAAAGGGGGCCAGCAGAGCTTCA-----	2726
OY	461	LeuMetProSerSerValAlaGlyGluThrHisSerValLeuAlaValProSerTrpArgsp	480
Db	2727	-----TAAAGCGCAAGGCTCAAGTGGCAGTCCCTGGAGCAGCAGCAT	2768
OY	481	HisSerValGluProLeuArgspProAsnProSerAspLeuLeuGluAsnLeuAspasp	500
Db	2769	CATTGACATGTCATCAGACAAATGC-----	2795
OY	501	serValPheSerLysArgHnLysLysLysLeuGluLeuAspLysArg-----ArgLy	518
Db	2796	-----CATTTCAACCTGTGAGCTGATGAAGCAGCGGCTCATCAAGAA	2837
OY	518	saArgTrpAspLleGlnArgLjleArgGlu-----GlnArgLleLeuGlnLnuArgLeuGlnLe	536

```

Db      2838 AAGGAGAGACTGTTCTCCTCGAGGAGGACACTGCGGAGCAAGGCGGAGCGGCTGCAGCC 2897
QY      536 uArMet---TyltLysLysGLyIleGInGInuSerGluProGluValThrSerPheP 555
Db      2898 TGAAGAGCCCGAGGAGMAAGAGAGAGGCTGCAGGAGCTGCCTGAGAGATCGAGGTCTGCC 2957
QY      555 eProGluProAspAspValGlu-SerLueMetIleThrProPheLeuProValValAlap 575
Db      2958 AGCCAACTTGTACTACATCAATGACGCGCATACCGACAGTGCAGGAGCAACCATCGTGCAGCT 3017
QY      575 heGlyLeuProLeuProLysLeuThrProGIn-----AsnPheGluLeuP 590
Db      3018 GGAAGAGAGCAACAGGAGAGCTGAGTACTCCACAGACATCGTGTGATCAAGCTTCGTCTC 3077
QY      590 rotPLeuAspGluArSer 596
Db      3078 CTTGGCTGAAGCCCGGCTCC 3097

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:      38.3      Length:      4403765
Score:          212.00    Matches:      159
Percent Similarity: 35.96% Conservative: 78
Best Local Similarity: 24.13% Mismatches: 240
Query Match:      6.55%   Indels:      184
DB:              4       Gaps:      29

US-10-054-935-2 (1-614) x US-09-103-840A-2 (1-4403765)
QY      10 AAlaAlaAlaProAlaGlyGlyAsnProGInGInArGLeuAspTyrGluAlaLa 29
Db      337754 GCCACTGCACACCCCTGCAGCAACAATTCCTCAACGGAT-----CAACGAGCC--- 337707
QY      30 AlaLeuLylGlyProGInuSerGluProGlyAlaAlaGluAlaInuAsPheLeuProArGInis 49
Db      337706 -----CACCCAAAGCCCTCCACCGGGGAGACCGCTGATCGGCAACAGCGGCCAA 337662
QY      50 ArGlyLeuLysGluProGlyProPheLeuAlaSerSerGInGlyLysSer----- 66
Db      337661 CGGCA-CTCCGGGGGACCGGGGGCGGACGCGGGGCCCGGGGTGGTTTCGCAACGCGC 337603
QY      67 -----ProAlaProSer-----ProAlaGlyCys 74
Db      337602 GCAACGCGCGCCACGAGCGACCGCGCGCGAGAGCGGCGATCCGCGGGGCGC 337543
QY      75 gLyGlyLysGlyArGdYLeuLeuLeuProAlaGlyAlaAlaProGlyGInGInGlu 94
Db      337542 GCGGAATCCTGTGCAGATATTGGCGGCGACCGCGGGGTCCGGCGGCGATCGGAACGAGCGGCC 337483

```



```

Db      2511 AGGACTAAACCCACCATGCTGCTGGGCTGAGTGTGCGGACGACACTACATC 2570
Qy      404 -----GlnLysGlyProSerThrHisProLysGluLysAlaPheSerGluIle 420
Db      2571 TGAGGCTGAATCGGGGCCGCTCTGCTCCACATCGTCGCC-----AGTGAACCG 2624
Qy      421 GlnAspLeuProTyrLeuSerThrThrGluMetTyrLeuCysArgTrpHisGlnPro 440
Db      2625 CAAATACACCACTTCTTGGGGGACCATCTCGGCCACACTGTCAATGCAACCCGCTCTGC 2684
Qy      441 ProSerProLeuProLeuArgLysSerProLysLysGluLutThrValAlaArgCys 460
Db      2685 CCAAGAGAGTCCCAAGAGAGGGGCCACACGAGAGCTTCAG-----2726
Qy      461 LeuMetProSerSerValAlaGluLutThrSerValLeuAlaValProSerTrpArgAsp 480
Db      2727 -----TAAAGCGGCAGAGCTCAAGTGGACATGCTCCCTGGAGCGACGAT 2768
Qy      481 HisSerValGluProLeuArgAspProAsnProSerAspLeuGluAsnLeuAspAsp 500
Db      2769 GATTGACATCGTCATGACGAGATGAC-----2795
Qy      501 SerValPheSerTyrArgHisAlaLys-IleGluLeuAspGluLysArg-----ArgLys 518
Db      2796 -----CATTCGCAACCTGGAGGCTGACATGAGCGGCTCAACAGAA 2837
Qy      518 SARGTTPASPILEGIAArgLysIle-----GlnArgIleLeuGlnArgLeuGlnLe 536
Db      2838 AAGGAGAGAGCTTCTCTCTGCAAGGAGCAGTCGCGAGAGAGATCGAGTGCCTGCGC 2897
Qy      536 uArgMet---TyrLysLysLysGlyIleGlnLysSerGluProGluValThrSerPhe 555
Db      2898 TGAGAGCCCCGAGAGAGAGAGGGGCTGCGAGAGCTGGCGAGAGATCGAGTGCCTGCGC 2957
Qy      555 eProGluProAspAspValGlu-SerLeuMetIleThrProPheLeuProValAlaLap 575
Db      2958 AGCCAACTGATGATCATCATGACGCGCATCACCGCTGCCAGGCCACCATCGTCAGCT 3017
Qy      575 heGlyArgProLeuProLysLeuThrProGln-----AsnPheGluLeuP 590
Db      3018 GGGAGAGACCCAGAGAGAGCTGGAGCTCCACAGACATCCGTGTCATCGCTCGCTC 3077
Qy      590 rOTrPLeuAspGluArgSer 596
Db      3078 CTTGGCTGAAGCCCGCTCC 3097

RESULT 12
US-09-718-815-1
; Sequence 1, Application US/09718815
; Patent No. 6455293
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6455293el motor proteins and methods for
; FILE REFERENCE: 1052
; CURRENT APPLICATION NUMBER: US/09/718,815
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4911
; TYPE: DNA
; ORGANISM: Human
US-09-718-815-1

Alignment Scores:
Pred. No.: 0.0216 Length: 4911
Score: 212.50 Matches: 181
Percent Similarity: 34.05% Conservative: 88
Best Local Similarity: 22.91% Mismatches: 261
Query Match: 6.56% Indels: 260

```

```

Db:      4      Gaps:      36
US-10-054-935-2 (1-614) x US-09-718-815-1 (1-4911)
Qy      14 ProAlaGly---GlyAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
Db      897 CTTGCTGGCTTGGCGAATGTGATGACGCC-----CTTAGG 932
Qy      33 GlyProGluAspGluProGlyAla-----AlaGluAlaHisPheLeuPro 47
Db      933 GGACACAGACAGAAAGGTGTGACCTTCCCTACAGGAGCTCCAGACTCCAGCTCCTGCTCCT 992
Qy      48 -----ArgHisArgLysLeuLysGluPro-----55
Db      993 CAGAGATTGCTGGGGGGCAACAGCCAGACCATCATGATGATGCTGTGTAGGCCCTCAG 1052
Qy      56 -----GlyProProLeuAlaSerSerGlnGlyLysProAlaProSerPro 71
Db      1053 CCGAGATTTCATGAGAGACCT---CAACACACTCAATATATCCAAATCGGGCCGCAACAT 1109
Qy      72 AlaGlyCysGlyLysGlyArgGlyLeuLeuProAla-----85
Db      1110 CAAGAACAGGTGTGTAGTACAGCAGACAGACAGCCAGCAATCATGCTACTGGGGC 1169
Qy      86 -----GlyAlaAlaProGlyLysGlnGlnGlnLysSer 95
Db      1170 TGAGATTGCTCGGCTGACATGAGATGATGATGATTAAGCGGCGAGAGTATAG 1229
Qy      96 TrpGlyGly-----SerValProLeuProCysProPro-----106
Db      1230 AGAGGATGGCGCTGAGGGCTATGATGATGATCTCCAGAGAAATGCAATGCTACAGAGA 1289
Qy      107 -----ProAlaThrLysGlnAlaGlyIleGlyLysLeuProAlaAlaAlaGlyAla 123
Db      1290 GAATGGGGCCCTCGGCT-----GCGGGTGAAGACCATGCGAGAGGCGCAT 1334
Qy      124 GlyCysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrLysSerLeuVal 143
Db      1335 CGATGCCATCAACACACCGCTGCTACCCAGCTCATGAGGCA-----GGAGGCCAACCT 1385
Qy      144 AlaAlaAlaLys-----GluProThrProTrp 152
Db      1386 GCTGTAGTACCAAGCGCGCGATGCAATGAGGCCATGTCGCTCATTCACAGACTACAT 1445
Qy      153 AlaGlyAspLysGlyAlaAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyPro 172
Db      1446 CCGGAGATGAGAGAGAGCTACGACTTAAGCTTTAGAGATGAAGCCATGAACAGATCCCT 1505
Qy      173 ProProLeuProLeuProGly-----179
Db      1506 GCGCGGAGCTCTGACAGGCGCTCGGCTAGAGCCCTACTCCTGGGTGCTTCTCCAGC 1565
Qy      180 ProProProLeuAla-----Pro 185
Db      1566 CGCCCGGCTTGGGGGAGCCCTGCGACCTCATGAGAGATGCTCGAGAGTATCCG 1625
Qy      186 ThrAlaThrAlaGlyThrLeuAlaAlaSer-----195
Db      1626 CAGGCGCCAGACAGAGACTGAGCGGCTAAAGAAGAGAGAGTACAGCAGCGAGAGAG 1685
Qy      196 -----GluGlyArgTrpLysSerMetArgLysSerProLeuGly-----208
Db      1686 CCCCAGAGAGAGAGCTTCATAAAGAGAGGCAAAATCCAAACAGAGAGAGAGAGAGAG 1745
Qy      209 -----GlyGlyGlyGlySerGlyAlaSer-SerGlnAlaLysLeuLys-- 223
Db      1746 GGATGAGAACAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Qy      224 -----GlnIleLeuLeuLysLeuLys 231
Db      1806 GGAAGGCGCGAGAGATGAAGATGAGAGACTCGGCGCATGAAGAGAGAGCTGTGACATCAG 1865
Qy      231 pleuIleGluGlnGlnGln-----GlnIleLeuLeuLysLeuLys 239

```

Db 3078 CCTGGCTGAAGCCCGCTCC 3097
 RESULT 11
 US-09-718-852-1
 ; Sequence 1, Application US/09718852
 ; Patent No. 6426193
 ; GENERAL INFORMATION:
 ; APPLICANT: Berard, Christophe
 ; APPLICANT: Freedman, Richard
 ; TITLE OF INVENTION: No. 6426193el motor proteins and methods for
 ; FILE REFERENCE: 1052
 ; CURRENT APPLICATION NUMBER: US/09/718,852
 ; CURRENT FILING DATE: 2000-11-22
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 4911
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-718-852-1

Alignment Scores:
 Pred. No.: 0.0216 Length: 4911
 Score: 212.50 Matches: 181
 Percent Similarity: 34.05% Conservative: 88
 Best Local Similarity: 22.91% Mismatches: 261
 Query Match: 6.56% Indels: 260
 Gaps: 36

US-10-054-935-2 (1-614) x US-09-718-852-1 (1-4911)

Qy 14 ProAlaGly---GlyAsnProGluGlnIarPleuAspTyrGluArgAlaAlaLeuGly 32
 ||||| ||| :|||
 Db 897 CCTGCTGGCTTGGCAATGTGATCAGCC-----CTTAGG 932
 Qy 33 GlyProGluAspGluProGlyAla-----AlaGluAlaHisPheLeuPro 47
 ||||| ||| ||||| :|||
 Db 933 GGACCAGAGCAAGAAAGTGTGTGACAGTTCCTACAGGAGACACCAAGCTCAGCTCGCTCT 992
 Qy 48 -----ArgHisArgLysLeuLysGluPro----- 55
 ||||| ||| :|||
 Db 993 CCAGATTCGCTGGGGGCAACAGCCAGACCATCTATCGCTGTGTGAGCCCTCAGA 1052
 Qy 56 -----GlyProPleuAlaSerSergInglYserProAlaProSerPro 71
 ||||| ||| :|||
 Db 1053 CCGAGATTTCATGGAGACCT---CAACACACTCAATATGCAATCGGCCCGCACAT 1109
 Qy 72 AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuProAla----- 85
 ||||| ||| ||||| :|||
 Db 1110 CAAGACAAGAGTGTGTGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1169
 Qy 86 -----GlyAlaAlaProGlyGlnGlnGlnGlnSer 95
 ||||| ||| :|||
 Db 1170 TGAGATTGCTGCTGCAGATGAGAGCTGATGAGATTAAGCGGGGCAAGCGAGTGTAGG 1229
 Qy 96 TrpGlyGly-----SerValProLeuProCysProPro----- 106
 ||||| ||| :|||
 Db 1230 AGAGAGTGGCGCTGAGGCTTATGATGTGTTCGAGAGAGATGCAATGCAAGCAAGCAAGCAAG 1289
 Qy 107 -----ProAlaThrLysGlnAlaGlyLysGlyGlyGluProAlaAlaAlaGlyAla 123
 ||||| ||| ||||| :|||
 Db 1290 GAATGGGGCCTCGGGCT-----GGGGGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1334
 Qy 124 GlyCysSerProAlaProLysTyrGlnAlaValLeuProLysGlnThrGlySerLeuVal 143
 ||||| ||| :|||
 Db 1335 CGATGCCATCAACAACCGCGCTCACCAGCTCATGAGCA-----GGAGGCCAACCT 1385
 Qy 144 AlaAlaAlaLys-----GluProThrProTrp 152
 ||||| ||| :|||
 Db 1386 GCTGCTAGCCAAAGCGCGGCGATGCGAATGAGCCATTGGTGGCTGATTCAGAACTACAT 1445

Qy 153 AlaGlyAspLysGlyLysAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyPro 172
 ||||| ||| :|||
 Db 1446 CCGGAGATTCAGAGAGTACGAGCACTACACTTCTAGAGAGTGAAGCAAGCAAGCAAGCAAG 1505
 Qy 173 ProPleuProPleuProGly----- 179
 ||||| ||| :|||
 Db 1506 GCGCGGAGCCTCTCAGCGGCTCGCTAGAGAGCCCTACTCTCTGCTGCTTCTCAGC 1565
 Qy 180 ProProPleuAla-----Pro 185
 ||||| ||| :|||
 Db 1566 CGCCCCGCTTCGGGGGAGCCCTCCAGCTCATGAGAGATGCTCGAGGTGATCCG 1625
 Qy 186 ThrAlaThrAlaGlyThrLeuAlaAlaSer----- 195
 ||||| ||| :|||
 Db 1626 CAGGCCAAGCAGAGCACTCGAGCGGCTTAAGAAGAGAGAGTCAAGCGGAGAGAGAG 1685
 Qy 196 -----GluGlyArgTrpLysSerMetLysSerProLeuGly----- 208
 ||||| ||| :|||
 Db 1686 CCCCAG 1745
 Qy 209 -----GlyGlyGlyGlySerGlyAlaSer-SerGlnAlaLysLeuLys-- 223
 ||||| ||| :|||
 Db 1746 GATGAGAGACAGAGCGGAG 1805
 Qy 224 -----GlnLeuLeuLeuGlnLeuAs 231
 ||||| ||| :|||
 Db 1806 GGAAGGGCGGAGAGTGAAGATGAGAGACTCGGGCAGTAAAGAGAGCTGTGTGAGCTCA 1865
 Qy 231 PleuLeuGlnGlnGlnGln-----Gln 239
 ||||| ||| :|||
 Db 1866 CTCAGACCCCGAG 1925
 Qy 239 nLeuGlnAlaLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 259
 ||||| ||| :|||
 Db 1926 GATCAATTCAG 1985
 Qy 259 AArgLysLeuArgMetGluArgMetGlnLeuValLys-----LysAspAsnGln 276
 ||||| ||| :|||
 Db 1986 GCTCAAGACAGATGTAGAGAAAGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2045
 Qy 276 nLysGlnArgHisLysLeuPheGlnGlyTyrGlnLysGlnLysGlnLysGlnLysGlnLys 296
 ||||| ||| :|||
 Db 2046 GCTGAGAGCGAGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2105
 Qy 296 uSerGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 316
 ||||| ||| :|||
 Db 2106 GGCACAAAGATCAAG 2165
 Qy 316 oProLysProPheSerCysGlyArgSergLysGlyLysLysLysLysLysLysLysLys 336
 ||||| ||| :|||
 Db 2166 GAAGCTGAGAGCGGCGGAG 2223
 Qy 336 ySerThrLysArgLysThrProValLysLysLeuAlaProLysSerLysVal---Lys 355
 ||||| ||| :|||
 Db 2224 -----GAGAGGAG-----CTGAAAGCTCAAGCGGAGAGAGAGAGAGAGAGAGAG 2270
 Qy 355 sThrLysThrProLysHisSerProLysLysGlnLysLysLysLysLysLysLysLysLys 374
 ||||| ||| :|||
 Db 2271 GGCAGAGTGGCCCTGATGAAGCAGATCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2330
 Qy 374 uThrValCysLysArgGlnLeu----- 381
 ||||| ||| :|||
 Db 2331 GACCAAG 2390
 Qy 382 -----ArgSergLysLys-----ThrPr 387
 ||||| ||| :|||
 Db 2391 GATCGAGCTCTGAGAGTCCAG 2450
 Qy 387 oGlu-----LysProArgSerSerValAspThr----- 396
 ||||| ||| :|||
 Db 2451 GGAGGTTTCTGCACTGAGAGCGGCTGGCCAAAGCCCATGTCTGAGCGGGGTGGAGAGCGGCT 2510
 Qy 397 -----ProProArgLeu-----SerThr-Pro----- 403

QY	56	-----GlyProProLeuAlaSerSerGlnGlySerProAlaProSerPro	71
Db	1053	CCGAGATTTCATGGAGACCTT---CAACACACTCAAAATATGCCAATGGGCCCCACACAT	1109
QY	72	AlaGlycysGlyGlyGlyGlyArgGlyLeuLeuLeuProAla-----	85
Db	1110	CACGACACAGGTGGTAGTGAACACAGACACAGACCAAGCAGCAAAATCAAGTCAGTCGGGCC	1169
QY	86	-----GlyAlaAlaProGlyGlnGlnGlySer	95
Db	1170	TGAGATTGCTGGGCTGCAGATGAGACTGATGAGATGATAGAGCGGGCAAGCAGTGAATGG	1229
QY	96	TrpGlyGly-----SerValProLeuProCysProPro-----	106
Db	1230	AGAGGATGGCGCTGAGGCTTATGATGATCTCTCCGAGAGATACCATGTCTACAGAAAGA	1289
QY	107	-----ProAlaThrLysGlnAlaGlyLLeuGlyGlyGlyProAlaAlaAlaGlyAla	123
Db	1290	GAATGGGGCCCTGGCGCT-----GGGGTGAAGACCATGCGAGGCCAT	1334
QY	124	GlyCysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGlySerLeuVal	143
Db	1335	CGATGCCATCAACAACCCGCTCACCACCATCATGAGCA-----GGAGGCCAACCT	1385
QY	144	AlaAlaAlaLys-----GluProThrProTrp	152
Db	1386	GCTGTGATCCACAGGCCCGCGATGCAATGAGGCCATGTGGCTGAATCCAGAACATCAT	1445
QY	153	AlaGlyAspLysGlyGlyAlaAlaSerProAlaAlaThrAlaAlaSerAspProAlaGlyPro	172
Db	1446	CCGGAGATCGAGGAGCTACGGACCTTAAGCTTTAGAGAGTGAACCATGAACAGATCCCT	1505
QY	173	ProProLeuProLeuProGly-----	179
Db	1506	GCGCCGACGCTCTCACGGCGCTCGGCTAGGAGCCCTACTCCTGGGTGCTTCCAGC	1565
QY	180	ProProLeuAla-----Pro	185
Db	1566	CGCCCGGCTTCGGGGGCGACCCCTGCCAGCTCCATGAGAGATGCTTCGGAGGTGATCCG	1625
QY	186	ThrAlaThrAlaGlyThrLeuAlaAlaSer-----	195
Db	1626	CAGGGCCACAGGAGCACTCGAGCGGCTTAAAGAAAGAGAGTCAAGCGCGGAGGAAGAG	1685
QY	196	-----GluGlyArgTrpLysSerMetArgLysSerProLeuGly-----	208
Db	1686	CCCCGAGAGGAAGCCTTCAAAAAGAGGCGAAACTCCACAGGAGAAACACCGAGAGAC	1745
QY	209	-----GlyGlyGlyGlySerGlyLysAlaSer-SerGlnAlaAlaCysLeuLys--	223
Db	1746	GGATGAGAACAGAGCGCGAGAGGAGAGAGAGGAAGAACGACGACAGATGGCTGTGAGGAGA	1805
QY	224	-----GlnIleLeuLeuGlnLeuLys	231
Db	1806	GGAAGGGCGCGAGATGAGATGAGACTCGGCAGTGAAGAGAGCCTGGTGGACTCAGA	1865
QY	231	PleuIleGlnGlnGlnGln-----Gln	239
Db	1866	CTCAGACCCCGAGGAGGAAGGTGAACCTTCAGAGCGGCACTGGCCGACTGACTGTGA	1925
QY	239	nLeuGlnAlaLysGlnLysGlnIleGlnGlnLeuLysSerGlnArgAspThrLeuAla	259
Db	1926	GATCGAAATCAAGACGAAGCTGATCGACGACACTGGAGAACGCCAGCGGGCGTTGCAGAC	1985
QY	259	ArgIleGlnArgMetGlnArgArgMetGlnLeuValLys-----LysAspAsnGln	276
Db	1986	GCTCAGACACACGATGAGGAAGAGCTGATTCCTGCTCGAGAACAAAGATCCGAGACACACA	2045
QY	276	uLysGlnArgHisLysLeuPheGlnGlyTyrGlnThrGlnGlnArgGlnGlnGlyLeu	296
Db	2046	GCTGGAGCGCACCGTGTGCTGCAGAACCTCAGACCAATGGAGTGCTATATCTGAGGAGAA	2105

QY 296 uSerGIuylsIleIyLsLeuGIuCySgInProGIuLeuSerIuIhSrGIInIhLeuPr 316
 Db 2106 GGCCAAACAAAGACGACGACTGTGAAGAGAGCGTGGGAGATGAACCGGACGTGCA 2165
 QY 316 oProIySProPheSerCySgIyArGserGIyIySgIyHISySArGIyIySerProPheGI 336
 Db 2166 GAAGCTGACAGGCCCGCCCAAGAAAGACACGCCCGGCTGTTAAGAACCACTGCCCTAC-- 2223
 QY 336 ySerThrGIuArGIySThrProValIySylsLeuAlaProGIuPheSerIyVal---Ly 355
 Db 2224 -----GAGAGGGAG-----CTGAAGAAAGCTACAGGCCAGAGTGGGTGAGATGAAGAA 2270
 QY 335 sThIySThrProIySHISerProIleIySgIuIu---ProCysGIySerLeuSerGI 374
 Db 2271 GGCCAAAGTGGCCCTGTATGAAGACAGATATGCTGAGAGAGCAACAGCGGCGGCTAATGGA 2330
 QY 374 uThValIySylsArGIuIeU----- 381
 Db 2331 GACCAAGAGAACGGGAGATGCCACAGCTACAGAGAGACGACGCGGACAGAGTTTCA 2390
 QY 382 -----ArgSerGIuIu-----ThrPr 387
 Db 2391 GATCCGAGCTCTGGAGTCCAGAAAGCGGCACAGAGATGTGCTTGAAGAGAAAGACCA 2450
 QY 387 oGIu-----LysProHrgSerSerValSphr----- 396
 Db 2451 GGAGGTTTCTGCACTAGAGCGCCGTGCCCAAGCCCATGTGAGCGGGTGGCGAGGCGGTGC 2510
 QY 397 -----ProProArGIeU-----SerThr-Pro----- 403
 Db 2511 AGGACTAAAGCCACCCATGCTGAGACTCTGGGGCTGAGAGTGTGCGGCCACACTACTATC 2570
 QY 404 -----GluIySgIyProSerThrHISySArGIySylsAlaPheSerSer-GluIle 420
 Db 2571 TGAGGCTGAATCAGAGGGCGCGCTGTGTCTCAGATCGTGGCC-----AGTGAACCG 2624
 QY 421 GluAspIeuProIyTyrLeuSerThrThrGIuMetIyLeuCySArGIyPhISgInProPro 440
 Db 2625 CAATATCAACCACTCTTGGGGGAGCAATCTGCGCCCACTGATGCAATGGACCCGCTCTGC 2684
 QY 441 ProSerProIeuProIeuArGIuSerSerProIySgIuIuThrValAlaIyArGys 460
 Db 2685 CCGAAAGAAAGTTCAGAAAGAGGGGCCACGCCAAGGCTTCAG----- 2726
 QY 461 LeuMetProSerSerValAlaGIyIuThrSerValIeuAlaValProSerTrpArgASP 480
 Db 2727 -----TAAAGCGGCAAGGCTCAAGTGGCAATGCCCTGGAGCGACAGGAT 2768
 QY 481 HISerValGIuProLeuArGIuAspProAsnProSerAspIeuGIuAsnIeuAspASP 500
 Db 2769 CATTGACATTCCTCATGACAGAAATGAC----- 2795
 QY 501 SerValPheSerIyArGIyHISAlaIyLys-LeuGIuIeuAspGIuIyArG-----ArgIy 518
 Db 2796 -----CATTGTCAACCTGAGAGGCTGACACTTGAAGGCGCTCATCAAGAA 2837
 QY 518 SAIGTTPAspIleGIuArGIyIleArGIuIu-----GluArGIyLeuGIuArGIeGIuIle 536
 Db 2838 AAGGAGAGAGCTGTCTCTGTCAGAGAGCACTCGGAGGAAGGGGAGCGCGCTGCAGGC 2897
 QY 536 uArGmet---TyrIySylsIySgIyIleGIuIeuSerGIuProGIuValThrSerPhePh 555
 Db 2898 TGAGAGCCCCGAGAGAAAGAGAGGGGCTGACAGAGCTGGCTAGAGATCGAGGTGTGGC 2957
 QY 555 eProGIuProAspAspValGIuIu-SerIeuMetIleThrProPheIeuProValAlaIar 575
 Db 2958 AGCCAAACATTCATACATCAATGACGGCATCACCGACTGCCAGGCCACCATCGGACGT 3017
 QY 575 heGIyArGIyProIeuProIySylsIeuThrProGIuIu-----AsnPhGIuIeuP 590
 Db 3018 GGAGAGACCAAGAGAGAGCTGGAGCTCACAGACACATCCGTGTGCTATCAGCTCTGCTC 3077
 QY 590 roTrPLeuAspGIuArGser 596

QY 465 SerValAlaGlyLeuThrSerValLeuAlaValProSerTrpArgAspHisSerValGlu 484
Db 28766 GACGTACCGCGCGGAGAGACATCCGCCATCTCGGGCGGAGAGAGCGCGCGGAG 28825
QY 485 ProLeuArgAspProAsnProSerAspLeuLeuGluAsn-----LeuAsp----- 499
Db 28826 GTCCTC-----GACGACCTGCTCCGGGACACTCCGCTGACCCCTTCGTC 28870
QY 500 -----AspSerValPheSerLysArg 506
Db 28871 CTCTACTCTCGAAGCGCGGGGTCTGGGCGACGCGACGCGAGCGGCTCTACGCGGCGCC 28930
QY 507 HisAlaValLeuGlu-----LeuAspGluLysArgArg----- 519
Db 28931 AACGCCACCTCGACGCGCTCCGCCCGCGCGCGCGCGCGCGCGAGACGCGACTCG 28990
QY 520 -----TrpAspLeu-----GlnArgIleArgGluArgIleLeu 531
Db 28991 GTCGCTGGGCGCTCTGGGCGCGGCGACGCGATGGCGCGCGCGCGCGCTACTGG 29050
QY 532 GlnArgLeuGlnLeuArgMetLysLysLysGlyIleGlnGlnSerGluProGluVal 551
Db 29051 CAGCGT-----CGCGGCACTCCGTC-----GATGAGCGCC----- 29079
QY 552 ThrSerPhePheProGluProAspAspValGlnSerLeuMetIleThrProPheLeuPro 571
Db 29079 ----- 29079
QY 572 ValValAlaPheGlyArgProLeuProLysLeuThrProGlnAsnPheGlnLeuProTrp 591
Db 29080 -----CGACCGCGGCGCTGGAGCACTGGCCCAAGCGCCCTGAG-----CCACGA 29121
QY 592 LeuAsp-----GluArgSerArgCysArgLeu 600
Db 29122 CGAGACCTTCGTCGCGCGGCGGCGGATGTCGACTG 29154
RESULT 5
US-08-781-802-7/c
Sequence 7, Application US/08781802
Patent No. 5969121
GENERAL INFORMATION:
APPLICANT: ALLEN, Larry
APPLICANT: AIKENS, John
APPLICANT: FOSTER, Michael
APPLICANT: VONSTEIN, Veronika
APPLICANT: DEMIRJAN, David
APPLICANT: CASADABAN, Malcolm
TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 S. Wacker Drive 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,802
FILING DATE: 10-JAN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/694,078
FILING DATE: 07-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/019,580
FILING DATE: 12-JUN-1996
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/009,704
FILING DATE: 11-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,995
FILING DATE: 01-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 95,963-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3147 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 245..1231
OTHER INFORMATION: /note="TSPA E101 sequence longest
OTHER INFORMATION: open reading frame; other possible start codons are TTG/Leu13, TTG/Leu15, GTG/Val143"
OTHER INFORMATION: TTG/Leu13, TTG/Leu15, GTG/Val143"
NAME/KEY: mat_peptide
LOCATION: 245..1231
US-08-781-802-7
Alignment Scores:
Pred. No.: 0.00939 Length: 3147
Score: 215.50 Matches: 157
Percent Similarity: 30.04% Conservative: 53
Best Local Similarity: 22.46% Mismatches: 193
Query Match: 6.66% Indels: 298
Gaps: 29
US-10-054-935-2 (1-614) x US-08-781-802-7 (1-3147)
QY 4 ArgSerAlaValPheLysAlaAlaAlaPro-----AlaGly 16
Db 2643 AAGAGCCCTTAAGAGAGGCGCGCTCCGCCGCTCCGCCGCGCGGCGGCGGAG 2584
QY 17 GlnAsnProGlnGlnArg-----LeuAspTyrGlnArgAlaAlaAla 31
Db 2583 GCCTCTCGAAGAGCGGCGCGCTCCCAAGCGCGGCGGCGGCGGCTTCG 2524
QY 32 GlyGlyProGlu-----AspGluPro 38
Db 2523 GGAAGCGCGAGATGAGCGCGGCGGCGGCGGCGGCGGCTTCGCTTCG 2464
QY 39 GlyAlaAlaGlnAlaHisPheLeu-----ProArgHisArgLysLeu 52
Db 2463 GCCTCCAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2404
QY 53 -LysGluProGlyProLeuAlaAspSerGlnGly---GlySerProAlaProSerPr 71
Db 2403 GCCCCTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2344
QY 71 AlaGlyLysGlyGlyLysGlyArgGlyLeuLeuProAlaGlyAlaAlaProGlyG1 91
Db 2343 TTCAGGCGGAGAGACTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2284
QY 91 nGlnGlnGlnSerTrpGlyGlySerValProLeuProCysProProAlaThrLysG1 111
Db 2283 TACTGCGCTCGCGGAGGAGGAGCTGGGCGGAGGCTCCAGGCGGCGGCGGCGG 2232
QY 111 nAlaGlyLysGlyGlyGlyGlyProAlaAla-----Ala-G 122
Db 2231 -CCAGGCGGAG 2173
QY 122 LysAlaGlyCysSerProArg-ProLysTyrGlnAlaValLeuProIleGlnThrGlySer 141

```

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-19

Alignment Scores:
Pred. No.: 0.121 Length: 38506
Score: 217.00 Matches: 177
Percent Similarity: 31.13% Conservative: 63
Best Local Similarity: 22.96% Mismatches: 256
Query Match: 6.70% Indels: 282
DB: 3 Gaps: 36

US-10-054-935-2 (1-614) x US-09-320-878-19 (1-38506)
QY 1 MetTherMetarSerAlaValAlaAlaAlaProAlaGlyAsnProGlu 20
Db 27161 GTACACCTGGCGCAAGTCCATCGCCGCCGCGCAAGGCGCATGATCTCC 27220
QY 21 GlnArgLeuAspTyrGlnArgAlaAlaAlaLeuGlyGlyProGluAspGluProGly 39
Db 27221 CTCGCCCTCGAC-----GAGCGCGCGCTCTGAAGGACTGAGCACTTCAGCGACTC 27274
QY 40 ---AlaAlaGlnAlaHisPheLeuProArgHisArgIleuLeuGluProGlyPro 57
Db 27275 TCCGTGCGCGCCGCTCAACGG-CCCCACCGCCGCGCTGTCTCCGGCGACCCGACGAT 27333
QY 58 -----CysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGly 65
Db 27334 CGAAGAACTCGCCCGCACCTGCGAGCGCGACGGGCTCGTCCGCGGATATCCCGGTGCA 27393
QY 66 SerProAlaProSerProAlaGly-----ProLeuAlaSerSerGlnGlyGly 73
Db 27394 CTACGCTCTCCACAGCGCGAGCTCGAGATCATCGAAGAGAGCTGCGCAGGTCTCGC 27453
QY 74 -----CysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGly 90
Db 27454 CGGACTCGCCCGCAGGCTCGCGACGTGCGCTCTCT-CCACCTTCGAAGGACCTGGA 27512
QY 91 GlnGlnGlnGlu-----SerTrpGly 97
Db 27513 TCACCGAGCGCGGTGCTCGAGCGCACTACTGTAACCGCACTGCGCATCGCGTGGCT 27572
QY 98 GlySerValPro-----LeuProCysProProProAlaThrLysGlnAla 112
Db 27573 TCCGCCCGCGCGTGAGACCTTGCGGTGACGGCTTCACCCACTTCATGAGGTGACGG 27632
QY 113 GlyIleGlyGlyLysProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyrGln 132
Db 27633 -----CCACCCCGCTCTCAACCATGACCTCCCGAAGCACTGCA----- 27671
QY 133 AlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGlu----- 148
Db 27672 -----CGGCGCTCGGCACTCTC-----GCCCGAAGAGGAGGCGCAGGAGCTC 27716
QY 149 -----ProThrProTrpAlaGly---AspLysGlyGlyAlaAla 160
Db 27717 TGGTCACTTCATCGCGCGAAGCCTTGCGCAACGCGCCACCATCATGATGCGGCCCTATCC 27776
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProPro----- 173
Db 27777 TCCCGACCGCAACCGCGCGACCGAGCTCCCGCACTACGCGCTTCAGACCGAGCGCT 27836

```

```

QY 174 -----ProLeuProLeuProGlyProProProLeuAlaProThrAlaThr 188
Db 27837 TCTGGCTGCAGACCTCCGCCCA-----CCAGCGCCCGCCAGCACTGGCGTTACCGGTG 27893
QY 189 AlaGlyThrLeuAlaAlaSerGlnGlyArgTrpLysSerMetArgLysSerProLeuGly 208
Db 27894 AGTGA-----AGCCGTGACGGCTCCGCCAGCGGACCTGTCCGGCGGT 27941
QY 209 GlyGlyGlyLysSerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnIleLeuLeu 228
Db 27942 GGATCGTCGCGTGGAGAGGACAGAACCCAGCTGTG----- 27982
QY 229 GlnLeuAspLeuIleGlnGlnGlnGlnGlnAlaAlaLysGluIleGlu 248
Db 27983 -----GCGCGCTGAGCGCGCGCGGAGCGGAGGTGAC 28015
QY 249 GlnLeuLysSer-----GluArgAspThrLeuLeuAlaArgIleGluArgMet 264
Db 28016 GTACTGGAAAGCCGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 28074
QY 265 GluArgArgMetGlnLeuValLysLysAspAsnGlnLysGluArgHisLysLeuPheGln 284
Db 28075 GACGACCGCGAGCGCTTACCGCGGTGTCTGCTCTGACGAGCAGCTGTGCCACAGT 28134
QY 285 GlyTyrGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 304
Db 28135 CGCCTGGGTGCAGGACACTGCGGAGCGGAGATCAAGCGCCCTGTGCTGCTCAACCA 28194
QY 305 GlnProGlnLeuSerGlnThrSer----- 312
Db 28195 GGGCGCGGTCTCGTGGAGCTGTGCACACCCCGCGACCCCGGACCGGCGCATGCTCTG 28254
QY 313 -----GlnThrLeu-----Pro 316
Db 28255 GGGCTTGGCGCGGTGCTGCTGCTTGAAGACCCCGAAGCGTGGCGCGCTGCTGACCT 28314
QY 317 ProLysProPheSerCys-----GlyArgSerGlyGlyGln 329
Db 28315 CCCGCGCACCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28374
QY 330 LysArg-----LysSerProPheGlySerThrGluArgGlyThrProValLysLys 346
Db 28375 CGGCGAGGACAGATGCGCATTCGCGCACCGAGGACTTCAGCGCGCGCTCCCGCGCG 28434
QY 347 LeuAlaProGluPheSer---LysValLysThrLysThrProLysHisSerProIleLys 365
Db 28435 ACCCTTCACGAGAGTGGCGCCACCGCGGCTGGACGCGCGCGCGCGCGCGCTCT 28488
QY 366 GlnGlnProCysGlySerLeuSerGluThrValLysLysArgGlnLeuArgSerGlnGlu 385
Db 28489 -----CATCACCGCGCGCACCGG 28506
QY 386 ThrProGlnLysProArgSerSerValAspThrProProArgLeuSerThrPro----- 403
Db 28507 AGCCCTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28566
QY 404 -----GlnLysGlyProSerThrHisProLysGlnLysAlaPheSerSer 418
Db 28567 CGTCAGCGGAGGCGGCGGAGCAACAGCCCGGAGCACCTCACTACCGC----- 28614
QY 419 GluIleGlnAspLeuProTyrLeuSerThrGluMetThrLeuCysArgTrpHisGln 438
Db 28615 -----CGAATCATCCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28649
QY 439 ProPro-----ProSerProLeu 444
Db 28650 CGCGCTCGAGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 28706
QY 445 ProLeuArgLysSerProLysLysGlnGlnThrValAlaArgCysLeuMetProSer 464
Db 28707 CCGAGAGCGCGCTTACCGCGCGGTGCTGCACACCGCGCGGCGGCGGCGGAT-CCGCTG 28765

```



```

? INFORMATION FORSEQ ID NO
? SEQUENCE CHARACTERISTIC
? LENGTH: 9551 base pair
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1507..1644
? FEATURE:
? NAME/KEY: Intron
? LOCATION: 1645..2511
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2512..8070
?
IS-08-056-Z00-.93

```

Alignment Scores:	
Pred. No.:	0.00229
Score:	238.00
Best Similarity:	35.6%
Best Local Similarity:	23.66%
Query Match:	7.35%
OB:	1
length:	9551
Matches:	162
Conservative:	82
Mismatches:	265
Indels:	176
Gaps:	30

US-10-054-935-2 (1-614) X US-08-056-200-93 (1-9551)

OY	12	AlAlaIaProAlaIaGlyGlyAlaAsnProGluGlnIaProIaGlyLeuAspIaGlyIaAlaIaAlaIaLeu	31
Db	4623	GGCCCAAGAGCAGAGAGCAAAAGAGCGGCCGA-----GAGTGAGCTGCATATGGCA	4673
OY	32	GlyGlyProGluAspGluProGlyIaAlaIaGluIaHisPheLeuPro---ArgHisArg	50
Db	4674	GGAGGAGAACGGGGCTCACCGCAGCAGACAGAGAGAGACAGAGCGCGCGCACTTCACATG	4733
OY	51	LysLeu-----LysGluProGlyPro-----Pro	58
Db	4734	GCACGTGGCAGCGGAGCAAAAGACGAGAGGGCGCTGCAGAGCTGTCCGCGACAGCCCC	4793
OY	59	LeuAlaSerGlnGly-----GlySerProAlaPro-----	69
Db	4794	ATTCCGGGAGCAGCGGGAGAGGACGACGTACGGCCGAGAGAGGCCACGACGCGGGAAACAAG	4853
OY	70	--SerProAlaGlyCysGlyGlyIaGlyIaArgGlyLeuLeuProAlaGlyIaAlaIa	88
Db	4854	GTTCCTCCC-----GGAGGAGAGAGCAAGAACAGCAGCGCGCCCGCAGCAGCGGA	4904
OY	89	ProGlyGlnGlnGluGluSerTyrGlyGlySerAlaProleuPro-----Cys	104
Db	4905	GAGGAGAGAAAGCTGCAGTTCCTGGAGGAAAGAGACGACTCCAGCGCGCGGAGCGTGC	4964
OY	105	ProProProAlaThrIaGlnAlaGlyIleGlyGlyGluProAlaAlaIaGlyIa	122
Db	4965	CCA-----ACAGCTCCAGAGAGAGAGAGACGGCTCCAGGAGGATCAGAGAGAGAG	5015
OY	123	--AlaGlyCysSerProArgProLysTyrGlnAlaValLeu-----	135
Db	5016	GCAGCAGGAGCAGCGCGCGCACCAAAATGAGGTGGCACTAAGCAAGCAAAAGAAAGAG	5075
OY	136	-----ProIleGlnThrGlySerLeuValAlaAlaAlaIaGluProThrPro	151
Db	5076	ACGGCCGCACACGCTGTACGCCAACGCCACGCTTCAAGAGCAGCTGAGGAAAGAAACAGCA	5135
OY	152	TrpAlaGlyAspLysGlyIaAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyIa	171
Db	5136	GCTCCTCGAGAGGAGGAGGAGAGCTACAGAGAGAGAGCAGCGCAAGAAAGAAAGAGCG---	5192
OY	172	ProProProIleuProIleuProGlyIaProPro--ProLeuAlaIaProThrAlaThrAlaGlyIa	190
Db	5193	-----CCAGAACAGGAGAGACAATACCGCGGAGGAAAGACAGCTGCAGCAGGA	5240

QY	191	ThrlreAlaAlaIasSerGIuGlnGlyArgTrpIyrsSerMetArgLysSerProLeuGlyGly	210
Db	5241	GGAGAGCAGCTGCTGAG---AGAGGAACGGGAGAAAAAGAACGCCA---GGAGCGGGA	5294
QY	211	GlyIySerGIyAlaIasSerSerGlnAlaIaCysLeuYsglnIleuLeuGlnIleu	230
Db	5295	AAGCAATATTCGGAAGCATATGAACCTGCACAG---AAGGAAGACAGCTCTGGAGAG	5352
QY	231	AspleuIleGlnGlnGlnGlnGlnIleuGlnAlaIalysGIuLysGIuIleGlnIleu	250
Db	5353	GAACCGGAGGAAGAGAACGGCCGAGAGCGGGAGAAAAAATATCCCGAGGAAGAGAGTTG	5412
QY	251	LysSerGIuArgAspTrhIleuLeuAlaArgIleGIuArgMetGIuArgTrMet-----	268
Db	5413	CAGCAGGAGGAAGACAGCTGCTG---AGACAGAACGGGAAGAGAGAGGCCAGAG	5469
QY	269	-----GlnIleuValLys	272
Db	5470	TGGGAGAGCAGTACCCGAAAAAGACAGCTGCAGCAGAAAGAACAGCACTGCTGAGA	5529
QY	273	LysAspAsnGIuLysGIuArgTrhIleLysLeuPheGlnGIuTrhGIuGlnIuArgIu	292
Db	5530	GAGGAACGGGAGAAAAAGAGA-----CTCCAGAGCGGGAGAGCAATATCGGGAG	5580
QY	293	GIuTrhTrhIleuSerGIuLysIleLysLeuGIuLysGlnProGlnIleuSerGIuTrhSer	312
Db	5581	GAAGAGGAGCTGCAGACAG-----GAGAAAGACAGCTGCTGGAGAGAACGG	5628
QY	313	GlnTrhIleuProProLysProPheSerCysGIuArgSerGIuLysGIuLysArgLys	332
Db	5629	GAGACG-----AGAAAGCGC	5643
QY	333	SerProPheGIySerTrhGIuArgLysTrhProValLysLysLeuAlaProGlnPheSer	352
Db	5644	CAGAGAGCTGGAGAGCAATATCGAAGAA-----GAGAGCTGCAGCAGAGAGAGAG	5697
QY	353	LysValLysTrhTrhLysTrhPro-----LysHisSerProIleLysGIuGIuProCysGIy	370
Db	5698	CAGCTGCGAGAGAGAACCGCAGAGAAAGAGAGCGCCAGGAGCGGAGAGCAATGTGCG	5757
QY	371	SerIleuSerGIuTrhValCysLysArgLysLeuArgSerGlnGIuTrhProGIuLysPro	390
Db	5758	GAGGAGAGAGAGCTGCACAGCAGAGAGAAAGACAGCTCTGTAAGAGAGAAAGGAGAAAGGA	5817
QY	391	ArgSerSerValAspTrhProProArgPheSerTrhProGlnLysGIyProSerTrhHis	410
Db	5818	AGCGGC-----CAGGAGCTGCAGAGCAATAT	5844
QY	411	ProLysGIuLysAlaIaPheSerSerGIuIleGlnAspleuProTrpIleuSerTrhTrhGIu	430
Db	5845	CGGGAGGGAGGAAGAGCTTCACGCGCCAGAAAAAGAAACAGCATACCGGGATGAGAGATCAG	5904
QY	431	MetTrpIleuCysArgTrpTrhIleGIuProProProSerProLeuProLeuArgIuIleSer	450
Db	5905	CGCAGTGTATCGAATATGGCAGTGGAGAACCAAAAAAGAAATGCAGTTCTGTATTAACAAG	5964
QY	451	-----ProLysLysGIuGIuTrhValAlaArgCysLeuMetProSerSerValAla	467
Db	5965	GTTTACTGTCAAGGACAGAGAAATGAACAGTTCCGGCGAGTTGGAAAGATTCACAGGTGGCGC	6024
QY	468	GIyGIuTrhTrSer-----ValIleuAlaValProSerTrpArgAspHis	481
Db	6025	GACAGACATCCACAGCAAGATCTGCAGCAGCTGCTGGGTGACACAGCAAGAGAGATGCT	6084
QY	482	SerValGIuProLeuArgAspProAsnProSer-----AspleuIleu	495
Db	6085	GAGCAAGAGAGAGCGCTGTCAGACAGGCCAAACAGCATTTCCAGAGAGAAAGACAGCTG	6144
QY	496	GIuAsnIleuAspAspSerValPheSerLysArgHisAlaLysIleu-----	510
Db	6145	GAGCGAGAGAGCAAAAGCAAGCCAAAGAGGCGCAGACAGAAATGCTCCCAAGAGAAAAAGCAG	6204

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 20:15:46 ; Search time 80 Seconds
(without alignments)
2353.744 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTMR5A/FAKAAAPAGNPE.....RSRCRLTIQKQTHRTCRK 614

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10054935/runat_14072003_115025_24679/app.query.fasta.1.775
-DB=Issued_Patents_NA -QEXT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blasum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINDLEN=0 -MAXLEN=2000000000
-USER=US10054935 @CGN 1.1.40 @runat_14072003_115025_24679 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	7.4	9551	1	US-08-056-200-93
2	238	7.4	9551	1	US-08-800-644-93
3	217	6.7	36778	4	US-09-105-537-5
4	217	6.7	38506	3	US-09-320-878-19
5	215.5	6.7	3147	2	US-08-781-802-7
6	215.5	6.7	3147	4	US-08-694-078-7
7	215.5	6.7	3147	4	US-09-058-260-7
8	214.5	6.6	441529	4	US-09-103-840A-1
9	213.5	6.6	11958	4	US-09-134-246-8
10	212.5	6.6	4911	4	US-09-718-692-1
11	212.5	6.6	4911	4	US-09-718-852-1
12	212.5	6.6	4911	4	US-09-718-815-1

Result No.	Score	Query Match	Length	ID	Description
13	212	6.5	4403765	4	US-09-103-840A-2
14	211.5	6.5	2214	3	US-08-864-038A-1
15	211.5	6.5	3331	3	US-08-864-038A-2
16	211.5	6.5	3331	3	US-08-864-038A-4
17	210.5	6.5	33529	4	US-09-144-085-3
18	207	6.4	4403765	4	US-09-103-840A-2
19	206	6.4	3211	2	US-08-574-959A-8
20	206	6.4	3211	4	US-09-357-014-8
21	206	6.4	3901	2	US-08-574-959A-6
22	206	6.4	3901	4	US-09-357-014-6
23	206	6.4	13578	3	US-08-785-420-1
24	203	6.3	3283	4	US-09-651-656-16
25	203	6.3	3283	4	US-09-650-855-16
26	203	6.3	441529	4	US-09-103-840A-1
27	201.5	6.2	2313	4	US-09-370-838-157
28	201.5	6.2	3239	4	US-08-927-219-5
29	201	6.2	43280	2	US-08-804-227C-1
30	200	6.2	4689	4	US-09-105-537-34
31	199.5	6.2	68750	3	US-09-335-409-1
32	199.5	6.2	68750	4	US-09-568-102-1
33	199.5	6.2	68750	4	US-09-567-969-1
34	199.5	6.2	68750	4	US-09-568-480-1
35	199.5	6.2	68750	4	US-09-568-486-1
36	199.5	6.2	68750	4	US-09-568-472-1
37	199.5	6.2	68750	4	US-09-567-899-1
38	199	6.1	1771	2	US-08-533-669A-7
39	199	6.1	1771	2	US-08-511-872-1
40	199	6.1	1771	4	US-09-183-861-7
41	199	6.1	1771	4	US-09-022-765-7
42	196.5	6.1	774	3	US-08-956-307B-12
43	196.5	6.1	778	3	US-08-956-307B-11
44	195.5	6.0	3132	2	US-08-224-482-3
45	195.5	6.0	3132	3	US-09-205-921-1

ALIGNMENTS

RESULT 1
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinhert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502

RL J. Biol. Chem. 274:32997-33001(1999).
 RN (6)
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
 RX PubMed:10433269;
 RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
 Doan A., Aakalu V.K., Lananan A.A., Sheng M., Worley P.F.;
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
 postsynaptic density proteins.";
 RL Neuron 23:583-592(1999).
 RN [7]
 RP INTERACTION WITH SPTAN1.
 RX PubMed:1509355;
 RA Bockers T.M., Maneza M.G., Kreutz M.R., Bockmann J., Weise C.,
 Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
 RT "Synaptic scaffolding proteins in rat brain. Ankrylin repeats of the
 multidomain Shank protein family interact with the cytoskeletal
 protein alpha-fodrin.";
 RL J. Biol. Chem. 276:40104-40112(2001).
 RN [8]
 RP FUNCTION.
 RX PubMed:11498055;
 RA Sala C., Plech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
 RT "Regulation of dendritic spine morphology and synaptic function by
 Shank and Homer.";
 RL Neuron 31:115-130(2001).
 RN [9]
 RP REVIEW.
 RX PubMed:10806096;
 RA Sheng M., Kim E.;
 RT "The Shank family of scaffold proteins.";
 RL J. Cell Sci. 113:1851-1856(2000).
 CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
 density (PSD) of excitatory synapses that interconnects receptors
 of the postsynaptic membrane including NMDA-type and metabotropic
 glutamate receptors, and the actin-based cytoskeleton. May play a
 role in the structural and functional organization of the
 dendritic spine and synaptic junction. Overexpression promotes
 maturation of dendritic spines and the enlargement of spine heads
 via its ability to recruit Homer to postsynaptic sites, and
 enhances presynaptic function.
 CC -1- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
 SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
 DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
 the PDZ domain (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
 neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
 CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms; 1, (shown here); 2, 3,
 4/5 and 5; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,
 CA1 region hippocampus and molecular layer of cerebellum).
 CC -1- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
 to high levels at 3-4 weeks before dropping slightly in adulthood.
 CC Expressed in the cortex and the molecular layer of the cerebellum
 at postnatal day 7. Isoform 2 expression does not change during
 development of both cortex and cerebellum. Isoform 4 expression
 decreases significantly during development of cortex but not
 cerebellum.
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 ANK REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF102855; AAD04569.2; -;
 DR EMBL: AF131951; AAD29417.1; ALT_INIT.

DR EMBL: AF159046; AAD42975.1; -;
 DR EMBL: AF141904; AAF02498.1; ALT_INIT.
 DR HSSP: P00519; IABL.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001660; SAM.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00023; ank; 7.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00536; SAM; 1.
 DR Pfam: PF00018; SH3; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00248; ANK; 3.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00454; SAM; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50086; ANK_REPEAT; 3.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR PROSITE: PS50105; SAM_DOMAIN; 1.
 DR ANK repeat, SH3 domain; Repeat; Alternative splicing.
 KW ANK repeat, 195
 FT REPEAT 195 210 ANK 1.
 FT REPEAT 212 245 ANK 2.
 FT REPEAT 246 278 ANK 3.
 FT REPEAT 279 312 ANK 4.
 FT REPEAT 313 345 ANK 5.
 FT REPEAT 346 378 ANK 6.
 FT REPEAT 379 395 ANK 7.
 FT DOMAIN 554 613 SH3.
 FT DOMAIN 663 757 PDZ.
 FT DOMAIN 2104 2167 SAM.
 FT DOMAIN 929 932 POLY-PRO.
 FT DOMAIN 1010 1015 POLY-HIS.
 FT DOMAIN 1022 1027 POLY-HIS.
 FT DOMAIN 1194 1199 POLY-GLY.
 FT DOMAIN 1850 1860 POLY-PRO.
 FT VARSPPLIC 1 613 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 615 654 SOGROESRDKARLFRHYTVGYSDSPDAPSLINDG
 > MALNAVGGPGGALPQPPALSSMPALGPRRSVWY
 1Y (IN ISOFORM 2).
 FT VARSPPLIC 646 654 MISSING (IN ISOFORM 3).
 FT VARSPPLIC 797 804 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 1930 1943 ISEDSOTSLSKPS -> QYRVVSSPDGDF (IN
 ISOFORM 5).
 FT VARSPPLIC 1944 2167 MISSING (IN ISOFORM 5).
 FT CONFLICT 1141 1141 S -> T (IN REF. 1).
 FT CONFLICT 1174 1174 S -> N (IN REF. 2).
 FT CONFLICT 1246 1246 R -> K (IN REF. 1).
 FT CONFLICT 1323 1323 A -> T (IN REF. 1).
 FT CONFLICT 1331 1331 S -> D (IN REF. 1).
 FT CONFLICT 1726 1726 S -> N (IN REF. 2).
 SQ SEQUENCE 2167 AA; 226333 MW; 3F47BB5A7B18BA86 CRC64;
 Query Match 5.5%; Score 178.5; DB 1; Length 2167;
 Best Local Similarity 22.6%; Pred. No. 0.61;
 Matches 136; Conservative 48; Mismatches 211; Indels 207; Gaps 29;
 QY 16 GGNPQRLDYERAAALGGDEPDGAEAHFL-----PRHKKLEPCPPPLASSGGSP----- 67
 DB 988 GGGRRKSLYH--SAALPRANNHPPNNHNNHAPPPPPRPP--EMETGGSPDDPP 1044
 QY 68 -----APSPAGCGGGRGLLLPAGA-APQQQSEMSGVPLPCRP----- 106
 DB 1045 PRLALGPQSLGWMGGGSPSTSGAPSPSHSSSGSGPQAPALRTFQLPPLRAASAM 1104
 QY 107 -----PATKAGIGGP-----AAAGACSPR-----PKYQAVLPIQT 139
 DB 1105 YVPANSGRKRKPLVKQTKVEEPPQKGSIPASSTPALRSPRPAPGSEKKSIPPT 1164
 QY 140 GSLVA-----AAKEPTPMAAGDKGASPA-----ATSDAPGPPPLPLP--- 178
 DB 1165 IIRKAPSTSSSGRSSSGSSTEAEPPTQPDGAGGGGSSPSAPAPATSPVPSPSPVPTPASP 1224

```

OY 179 -GPEPLAPTAAGT-----LAASEGRL--KSMRKSPL-----GGGGSSGSSQAAC 221
D 1225 SGPATLPTSGFALVGAARREGGKMONERRRSTLLSLDAGDEGDSGLGFGGPPG- 1283
OY 222 LKQILLQLDLIEQQOOLQAKKEIEELKSERDTLLARIERMERRRQVLKDKKERHK 281
D 1284 -----PRLR-HSKSIDGMEFSAEPL-----RLSSGSSSGGYGVAAVAGSR 1322
OY 282 LFGCYETEEREELKSEKIKLECOPELSETSQILPPKPPSCGSGCKHKRRKSPG----S 337
D 1323 AYGSGSSS-----SAFTSELPPLPVLPGLGKALDPAKLPGLALAA 1363
OY 338 TER-----KTPVKKIAPEFSKVTKTPKHSR-----IKEPQGSLSERVCK 378
D 1364 RERLAKSSGCGGTPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1423
OY 379 RE-----LRSGE-----TPKPRSSV--DTPPR-----LSTPOKGPSTHPK----- 412
D 1424 RELGYRAGLGSQEKALVAPSPAPARRSLHLRLPPAPGVPLLLQLGPEPPTPHGVAKAW 1483
OY 413 EKAFSEIEDLPYLTSTEMYLCRMHQ-----PPSP---LPLRESSPKK 453
D 1484 RTAAPEPERLP-LHVRFLNCOARPPAGTSGSSTEDGPGVPPSPRRVLPSTSPRG 1542
OY 454 EE 455
D 1543 NE 1544

RESULT 12
REL_HUMAN STANDARD: PRT: 755 AA.
AC 092766;
DT 13-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RAS-responsive element binding protein 1 (RREB-1).
GN RREB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE=Thyroid carcinoma;
RX MEDLINE=96413283; PubMed=8616445;
RA Thiaagalingam A., de Bustros A., Borges M., Jasli R., Compton D.,
RA Diamond L., Mabry M., Ball D.W., Baylin S.B., Nelkin B.D.;
RT "RREB-1, a novel zinc finger protein, is involved in
RT differential response to Ras in human medullary thyroid
RT carcinomas."
RL Mol. Cell. Biol. 16:5335-5345(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE
CC DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE
CC PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE
CC OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL
CC DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY AND PANCREAS. NOT FOUND IN THE BRAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U26914; AAB19094.1; -
CC HSSP: P08045; 12NF.
CC TRANSFAC: T01975; -

```

```

DR Genew: HGNC:10449; RREB1.
DR MTM: 602209; -
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 4.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; znf_C2H2; 4.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 315 337 C2H2-TYPE.
FT ZN_FING 461 483 C2H2-TYPE.
FT ZN_FING 580 602 C2H2-TYPE.
FT ZN_FING 608 630 C2H2-TYPE.
SQ SEQUENCE 755 AA; 79865 MW; 28D863FE62FE8415 CRC64;

Query Match 5.4%; Score 174; DB 1; Length 755;
Best Local Similarity 19.9%; Pred. No. 0.36;
Matches 142; Conservative 80; Mismatches 215; Indels 276; Gaps 32;

OY 34 PEDERGAEAHFLPRHKKLEKPPPLASSGCGSPAPSPA-----GCGKGRGLLPAGAA 88
D 28 PEEEGSGSE-----QSP-----CPAPGPSLPVTLGPGILSEPMAPAPAA 68
OY 89 -PGQDEESWGSGVPLPCP-----PPATKQAGIGBPAAAGAGC--SPRPYQAVLP 136
D 69 TPPEPPAQLOGPVOLANPIYSSALVSSPPL-----VSSALLSGTALLRLPRKRPPLLP 123
OY 137 -----IQTGSLVAAA-----KEPTP-----WAGDK--GGAASPAATA 166
D 124 KPVTIELPLASTIAQIITSSVASAPTLKTKVADPGAPAGTSGNTTASDLSGGVPAKAAAT 183
OY 167 SDPA-----GPPPLPLGPPLPLATAA--GTLAASEGRKSMKSPPLGGGG----- 212
D 184 ATPAATTSKPESSEPPAPASSPEASPTEQCAPATSKRKGRKMRSPRANSGVDLDS 243
OY 213 -----SCA----- 215
D 244 SGERASTEKMLATDTTKFSPFLQTAEDNTQDEVAGAPADHNGSPDEQSGSPEDKLLRA 303
OY 216 --SSQAACLKQILLQLDLIEQQOOLQAKKEIEELKSERDTLLARIERMERRRQVLK 273
D 304 KRNSYTNCLQKITCPHCPRVFPWMASSIQRHXLFTHTDSQDAETFAAXGEVLD-----LTSR 359
OY 274 DNE-----KERKKLQGYETERETELSE----- 298
D 360 DREQPSGATELROVAGDAVEQATAETASPVHHEERKGSHEPEEHGTESTGADG 419
OY 299 -KILECOPE-----LSETQTLPPKPPSCGSGKGRKSPFGSTERKTPVKKLAPFSK 353
D 420 GRGKVE-QPEPGRLRHQAGL-----QAGGRRKRGPRGCAASGE--QKLACDTGC 467
OY 354 VKTKPRHSPKPECGSLSEYVCKRELRSQETPEKPRSSVDPPLRSTPOKGPSTHPKE 413
D 468 KSEFK-----LQTLER--HRAKARQEKDKGCGAT-----TAEEGSPAPAE 509
OY 414 KAPSEIEIDLPLYLTSTEMYLCRMHQPPSPPLPRLSESPKKEEYVARCLM-----PSSVAG 468
D 510 E-----EKPPETAEVESAPGAGEAPAEKLAETEGPSPD--G 545
OY 469 ETSVLAVSWRDSVEPLRDPNPDLLENLDD-----SVFSK-----RHAKELDE 514
D 546 ESAAEKSSSEKSDDDKPKTDSKPSVASKADKKRKYCVCKNRKWSIQLDTRHMRSTIGE 605
OY 515 KRRKRMIDR-----IREQRIQLQILRMV-----KKKGIOSEBEVT 552
D 606 RPYKQCQCECTFTLKSLVHRHRIHQAKARAKHHGKSDKEEGEEDSENEST 658

RESULT 13
CSP_PLAKU STANDARD: PRT: 351 AA.
ID CSP_PLAKU
AC P04922;

```

DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Circumsporozoite protein precursor (CS).
 OS Plasmodium knowlesi (strain nurli).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5852;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85272582; PubMed=4023712;
 RA Sharma S., Spec P., Mitchell G.H., Godson G.N.;
 RT "Diversity of circumsporozoite antigen genes from two strains of the
 RL malarial parasite Plasmodium knowlesi.";
 RL Science 229:779-782(1985).
 CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
 CC VERTEBRATE HOST).
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M11031; AAA29540.1; -;
 DR PIR: A26253; OZ20KU.
 DR InterPro: IPR003067; Circmsprzoite.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF00090; TSP_1; 1.
 DR PRINTS: PR01303; CIRCMSPRZOITE.
 DR SMART: SM00209; TSP1; 1.
 DR Malaria: Sporozoite; Repeat; Signal.
 FT SIGNAL 1 19 PROBABLE.
 FT CHAIN 20 351 CIRCUMSPOROZOITE PROTEIN.
 FT DOMAIN 101 235 15 X 9 AA TANDM REPEATS OF A-A-G-A-G-G-
 FT E-Q-P.
 SO SEQUENCE 351 AA; 34782 MW; A85E87A152E6485B CRC64;
 Query Match 5.3%; Score 173; DB 1; Length 351;
 Best Local Similarity 33.5%; Pred. No. 0.19;
 Matches 75; Conservative 12; Mismatches 95; Indels 42; Gaps 11;
 QY 16 GGNPEORDYERAAALGGPEDEPGAAEAHFLPRHRKLEPGPPPLASOGGSPAPSPACG 75
 DB 63 GKKPEGADKEK-----KEKEKEKEEPKPKNEKKLKPDEPAAGAGEGPA---AGAG 114
 QY 76 GK---GKGLLPAGAPAGQDESN-GGSVLPCCPPPTKQAGIGEPAAAGAGCSPPRK 130
 DB 115 GEOPAAAGAGEGEPAAAGAGEGEPAAAGAGEGEPAAAGAGEGEPAAAGAG-GEPPA 173
 QY 131 YQAVLPITGSLVA-AKEPTPMAGDKG-----GAASPAATA--SDPA-----GPPPLP 176
 DB 174 AGA-----GGQPPAAGAGEGEPAAAGAGEGEPAAAGAGEGEPAAAGAGEGEPAP 228
 QY 177 LPPPPPLAPTAGTLAGSEGNKMSRKSPGLGGGSGASSQA 220
 DB 229 APRRQPPAGAVAGDGA-----RGNGAGAGGCGGCONNGA 263
 RESULT 14
 SFPQ_HUMAN STANDARD: PRT; 707 AA.
 AC P23246; P30808;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
 DE binding protein-associated splicing factor) (PTB-associated splicing
 DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
 GN SFPQ OR PSF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal Brain;
 RX MEDLINE=93194059; PubMed=8449401;
 RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
 RT factor.";
 RL Genes Dev. 7:393-406(1993).
 RN [2]
 RP SEQUENCE OF 312-707 FROM N.A.
 RC TISSUE=Fetal skeletal muscle;
 RX MEDLINE=90091812; PubMed=2480877;
 RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
 RT "Cloning and characterization of a myoblast cell surface antigen
 RT defined by 24.1D5 monoclonal antibody.";
 RL Development 105:723-731(1989).
 RN [3]
 RP SEQUENCE OF 48-68 AND 213-246.
 RX MEDLINE=93176127; PubMed=8439294;
 RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 RT "Purification and characterization of a DNA-binding heterodimer of 52
 RT and 100 kDa from HeLa cells.";
 RL Biochem. J. 290:267-272(1993).
 CC -1- FUNCTION: ESSENTIAL. PRE-mRNA SPLICING FACTOR REQUIRED EARLY IN
 CC SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
 CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
 CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
 CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
 CC ECTOKINASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X70944; CAA50283.1; -;
 DR EMBL: X16850; CAA34747.1; -;
 DR PIR: A43557; A43557.
 DR PIR: S29770; S29770.
 DR HSSP: P11940; ICYT.
 DR SWISS-2DPAGE: P23246; HUMAN.
 DR Genew; HGNC:10774; SFPQ.
 DR MIM: 605199; -;
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 2.
 DR SMART: SM00360; RRM; 2.
 DR PROSITE: PS50102; RRM; 2.
 DR PROSITE: PS00307; RRM_RNP_1; 1.
 KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
 KW Alternative splicing.
 FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
 FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
 FT REPEAT 9 11 3 X 3 AA REPEATS OF R-G-G.
 FT REPEAT 19 21 1.
 FT REPEAT 25 27 2.
 FT REPEAT 25 27 3.

```

FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPLIC 663 707 RTERFGGAGPVGCGPRGMPGTPAGYGRGREGS
FT CONFLICT 243 243 G -> R (IN REF. 3).
SQ SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;

Query Match 5.38; Score 172.5; DB 1; Length 707;
Best Local Similarity 22.38; Pred. No. 0.39; 200; Indels 219; Gaps 32;
Matches 137; Conservative 58; Mismatches 200;

QY 49 HRRLKEPGRPLAS-----QGSPAPSPAGCGKRGILLPA-----GAAPGQ 91
D 94 HQQQPPPPQDSSKPVVAGGPAP-----GVGSAPPASSAPATPPISGAPG- 144
QY 92 QESWGSVLPCCPPRATKAGIGGEPAAAGCSPRKYQAVLPITGSLVAAAKEPT- 150
D 145 -----SGRPPTPPPAPVTSAPRGPAP-----PPSPSSGV-----PTT 177
QY 151 -PMAGDGAASAPATSDPAPRPLPGRP-----PLPTAGTGLAASGKWKSMRK 204
D 178 PPOAG--GPPPPPAVNGPGRGPGKQGPGRGKMGPGGKGGGLSTPG--GHPK 232
QY 205 SPLGGGSGASSQAACILQLDLLEQOQO-----LQAKEIEELKSBDT 256
D 233 PRHNGGGERPGRQ-----HHPRYHQNHQGPFGGPGGRSEK----- 271
QY 257 LLAIEEMERMOVLKQDNK---ERKLLFG---YETERRETELSEK----- 299
D 272 -ISDSEFGKANLSLRPRGERTYQRCRLFVGNLPALITEDEEFRLKAYGEPGEVINK 330
QY 300 -----TKLSCPELSTQTLPPKP-----FSCGRSGKHKRKSPP----- 334
D 331 GKGGFKLKSRLALAEAKELDDTPKRGQLRYRFTHAALSVNLSPVSNELLEEA 390
QY 335 ---FGSTER-----KTPYKLLAPFSK---VKTKPKSPKKEE 367
D 391 FSGGPIERAVVIYDGRSTGKGIYEFASKPAARKAFENGSEVFLTTTPR--PYIVE 448
QY 368 PCGSLSTVCKRELRSQETP--EKPNSSVTPPRLSTPQKGPSTHPEKAFSSIEDLPY 425
D 449 PLEQLDDEDLPEKLAQKPMYQKER---ETPPRFA--QHG-----TFEYE----- 489
QY 426 LSTTEWYLCRWHPPPPLPLRESSPKKEFTVARCLMPSSVAGETSVLAVPSMDHSEV 485
D 490 -----YSQWK-----LDEMEKQOREOVENMDADOKLESEN--EDAEHIOANL 534
QY 486 LRDPNPSDL-----LENLDVSFSKR--HAKLEDEKRRKMDIQTREORI--LQ 532
D 535 LR-----QDLMRQBELRMELHNOEQMKRKMQLRQEEERRRREEMIRQRMEDQMR 590
QY 533 RLQLRMKKKKGIOE 546
D 591 ROREESYSRMGYMD 604

```

RESULT 15

HRX_HUMAN

AC 003164; Q14845; Q16364; Q13743; Q13744; Q9UMAJ;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein HRX (ALL-1) (Tithorax-like protein).

STANDARD; PRT; 3969 AA.

Q03164; Q14845; Q16364; Q13743; Q13744; Q9UMAJ;

01-OCT-1993 (Rel. 27, Created)

01-NOV-1995 (Rel. 32, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)

Zinc finger protein HRX (ALL-1) (Tithorax-like protein).

```

GN MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93046667; PubMed=1423624;
RA Tkachuk D.C., Kohler S., Cleary M.L.;
RT "Involvement of a homolog of Drosophila tithorax by 11q23
RT chromosomal translocations in acute leukemias.";
RL Cell 71:691-700(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96290553; PubMed=8703835;
RA Nilsson I., Loechner K., Siegler G., Grell J., Beck J.D., Fey G.H.,
RA Marschalek R., Sieglar G., Grell J., Beck J.D., Fey G.H.,
RT "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT translocations to chromosomal region 11q23 and acute leukemias.";
RL Br. J. Haematol. 93:966-972(1996).
RN [3]
RP SEQUENCE OF 1-1909 FROM N.A.
RX MEDLINE=93390935; PubMed=8378076;
RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA Koeda Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT "Two distinct portions of LTR19/ENL at 19p13 are involved in t(11;19)
RT leukemia.";
RL Oncogene 8:2617-2625(1993).
RN [4]
RP SEQUENCE OF 1317-2328 FROM N.A.
RX TISSUE=Brain;
RL MEDLINE=93265134; PubMed=1303259;
RA Diabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT "A tithorax-like gene is interrupted by chromosome 11q23
RT translocations in acute leukemias.";
RL Nat. genet. 2:113-118(1992).
RN [5]
RP SEQUENCE OF 1251-1538 FROM N.A.
RX MEDLINE=94215165; PubMed=8162575;
RA Gu Y., Alder H., Nakamura T., Schlichman S.A., Pressad R., Caneani O.,
RA Saito H., Croce C.M., Caneani E.;
RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT involved in acute leukemia.";
RL Cancer Res. 54:2326-2330(1994).
RN [6]
RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX MEDLINE=95322025; PubMed=7598802;
RA Morigioli D., Burnett R., McCabe N., Thirman M., Gall H., Yu H.,
RA Rowley J.D., Diaz M.O.;
RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT ttx zinc-finger domain, and alternative splicing.";
RL DNA Cell Biol. 14:475-483(1995).
RN [7]
RP SEQUENCE OF 1212-1603 FROM N.A.
RX MEDLINE=95315013; PubMed=7794749;
RA Marschalek R., Grell J., Loechner K., Nilsson I., Siegler G.,
RA Zweckroner I., Beck J.D., Fey G.H.;
RT "Molecular analysis of the chromosomal breakpoint and fusion
RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT translocation t(4;11).";
RL Br. J. Haematol. 90:308-320(1995).
RN [8]
RP SEQUENCE OF 1421-1540 FROM N.A.
RX MEDLINE=94020842; PubMed=841518;
RA Forster A., Rabbits T.H.;
RT "A method for identifying genes within yeast artificial chromosomes:
RT application to isolation of MLL fusion cDNAs from acute leukemia
RT translocations.";
RL Oncogene 8:3157-3160(1993).
RN [9]
RP CHROMOSOMAL TRANSLOCATION WITH GAST.
RX MEDLINE=20183971; PubMed=10706619;
RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,

```



```
QY 331 -----RKSPPGSTERKTPVK-----KLAPESKVT-----KTPKH 361
Db 250 EDSLKIKIRTPSAIFQOATIKIKIRACKLSPLSKFKTKGLQIGRKGVQIYRRGRPPST 309
QY 362 SPIKEEPCSL--SEYVCKRELRSQETPEKPRSSVD-----TPPRLSTPOKGPSTH--- 410
Db 310 ERIK-TPSGLLINSELEKPOVKRDKEGTPTLKEDTVVRQSPRRIRIKPVRIIPSSKRTD 368
QY 411 -----PKEAFSSSEIEDLPYLS 427
Db 369 ATIAKOLLORAKKGAQKKIEKEAAQLGRKVTQVKNIROFIMPVSAISSRI-----IK 423
QY 428 TTEMYLGRMHQPPSPPLPLRESSPKKEETVARCLMSSVAGETSVLAVPSWRDHSVEPLR 487
Db 424 TPRRFIEDEDYDPPIKIARLESTPN-----SRESAPSCGSSSEKSSAASHSSOMSSDSR 478
QY 488 DPNPSDLLENLSDSVFSKR-----HAKLELDKRRRRMDIORIREQRIOR 533
Db 479 SSSPS--VDISTDSQASEIQLVPEERSDTPVEVHPPLPISQSPENESNDRSRYSYSER 536
QY 534 -LQLRMYKKKGIOESEB-EVTSFEPEDVESLMTPELPPVAFGRPLPKLTPQNF---E 588
Db 537 SFGSRRTTKLSTIQSAPQOQTSSSPPP-----PLL-----TPPPPLQPASSISDH 581
QY 589 LPWL 592
Db 582 TPWL 585
```

Search completed: July 14, 2003, 11:49:04
Job time : 29 secs

SUMMARIES

(without alignments)
4242.333 Million cell updates/sec

Scoring table:

BLOSUM62		
Xgapop	10.0 ,	Xgapext 0.5
Ygapop	10.0 ,	Ygapext 0.5
Fgapop	6.0 ,	Fgapext 7.0
Delop	6.0 ,	Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

32308132

02 or 5
0132

32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

MODEL=framed_p2n.model -DEV=xlh
-O/cgml2/USPO.spool/US10054935/runat.14072003.115025.24667/app_query.fasta.1.775
-DB=EST -OQM=fastat -SUFFIX=1st -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-DONTS=bits -START=1 -END=1 -MATRIX=bitsun62 -TRANS=humand40.cci -LIST=45
-DOTALIGN=100 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=5 -ALIGN=15 -MODE=LOCAL
-OUTFNM=p1c0 -MORF=excl -HEAPSIZE=500 -MINLEN=2000000000
-USER=US10054935 -CGCM=1.1.1906 -exrunat.14072003.115025.24667 -NCPU=6 -ICPU=3
-NO_MMAP -LAREGORY=NING -SCORES=0 -WAIT -OSBPLOC=100 -LONGIOS
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELETE=7

```

Database :

- 1: em_estaa:*
- 2: em_esthnm:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estvw:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_estl:*
- 10: gb_estri:*
- 11: gb_hic:*
- 12: gb_estr3:*
- 13: gb_estv:*
- 14: gb_estf5:*
- 15: em_estfun:*
- 16: em_estlom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_man:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match
---------------	-------	----------------

Description

1	3110	96.0	3161	11	AK014463	AK014463	Mus muscu
2	1715	53.0	1304	11	AK015496	AK015496	Mus muscu
3	1324	40.9	765	14	BQ014717	BQ014717	UI-H-ED1-
4	1302	40.2	870	14	BQ228745	BQ228745	AGENCYCOURT
5	1197	37.0	854	14	B1665707	B1665707	603287711
6	1115.5	34.5	736	13	B1661979	B1661979	603504574
7	1114.5	34.4	662	12	BF284375	BF284375	EST448966
8	1104	34.1	695	14	BQ178560	BQ178560	UI-M-EVO-
9	1097	33.9	686	14	BQ257208	BQ257208	NISC_Ko11
10	1077.5	33.3	891	14	BQ081177	BQ081177	AGENCYCOURT
11	1074.5	33.2	686	10	AW965141	AW965141	EST377214
12	1073	33.1	723	14	BQ187281	BQ187281	UI-E-ED1-
13	1069.5	33.0	811	13	B1154331	B1154331	602405281
14	1059.5	32.7	1087	12	BG421221	BG421221	602454171
15	1055.5	32.6	615	9	A1102093	A1102093	EST211382
16	1047	32.3	631	10	BE693636	BE693636	MR4-BT035
17	1036.5	32.0	698	13	B1905419	B1905419	603167481
18	1033	31.9	623	13	B1957952	B1957952	603198345
19	1029	31.8	637	10	BE693616	BE693616	MR4-BT035
20	1020	31.5	661	10	BB665412	BB665412	BB665412
21	1014	31.3	583	13	BM263966	BM263966	1933B05.x
22	997	30.8	761	13	B1694621	B1694621	603347843
23	989.5	30.6	815	13	B1694963	B1694963	603296616
24	981	30.3	549	9	A1800794	A1800794	wg135C04.x
25	978	30.2	701	10	BE287062	BE287062	601091037
26	971	30.0	551	13	B1667371	B1667371	603392145
27	961	29.7	600	13	B1987618	B1987618	3203-77 M
28	956.5	29.5	683	12	BF045095	BF045095	BP250018A
29	939.5	29.0	551	9	AA818434	AA818434	UI-R-AO-A
30	924	28.5	530	10	AW143159	AW143159	EST293455
31	918	28.4	541	13	EM264268	EM264268	1933B05.y
32	918	28.4	618	13	BM487396	BM487396	PM2m.n.PK0
33	885	27.3	651	12	BG480294	BG480294	602529340
34	885	27.3	699	10	AV762388	AV762388	AV762388
35	873	27.0	510	9	AA681060	AA681060	VUL31F05.x
36	871	26.9	702	9	AJ395334	AJ395334	AJ395334
37	867	26.6	519	10	AW049945	AW049945	UI-M-BH1-
38	860	26.6	985	14	BQ711779	BQ711779	AGENCYCOURT
39	852.5	26.3	512	9	A1170640	A1170640	EST216574
40	841	26.0	501	10	AW854130	AW854130	RC3-CT025
41	835	25.8	479	12	BF116257	BF116257	7n79D02.x
42	829	25.6	463	9	A1954139	A1954139	wx80B09.x
43	829	25.6	508	14	BQ127983	BQ127983	1179908.y
44	825	25.5	460	9	AT582361	AT582361	tg67D10.x
45	825	25.5	1075	12	BF134513	BF134513	601784919

ALIGNMENTS

LOCUS	AK014463	3161 bp	mRNA	linear	HTC_19-JAN-2002
DEFINITION	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:121402D02;PRICHEXTENSN domain containing protein, full insert sequence.				
ACCESSION	AK014463				

ACCESSION	AK014463
VERSION	AK014463.1
KEYWORDS	GI:12852332
SOURCE	HTC; CAP trapper.
	Mus musculus (strain:C57BL/6J) 16 days embryo head cDNA to mRNA.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL MEDLINE
99279253
PUBMED 10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE
20499374
PUBMED 11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE
20530913
PUBMED 11076861

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Alzawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Konno, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Niki, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabaldi, F., Suzuki, R., Tomita, M., Wagner, E. L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotamian, M., Hume, D. A., Kamly, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Momberts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyono, O., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, L., Wyshaw-Borls, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL MEDLINE
21085660
PUBMED 11217851

5 (bases 1 to 3161)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F., Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kuzihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunakawa, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Shinohe-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

COMMENT

Fax:81-45-503-09216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research at Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGACGAAGATCCAGAAGCTCTTTTCTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer abapcr of sequence [5' GAGAGGAGCTTCGGCTTAATTAAATTAATGCCCGCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size compressed longer than 7 kb was selected before cloning. Vector: a modified plasmidscript KS(+), after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

SOURCE

location/Qualifiers

1..3161

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:4121402D02"

/db_xref="MGD:MGI:1907076"

/db_xref="taxon:10090"

/clone="4121402D02"

/tissue_type="head"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="16 days embryo"

316..2166

/note="PRICEXTENSIN domain containing protein
data source: InterPro, source key:IPR002965, evidence:ISS putative"

/codon_start=1

/protein_id="BAB29369.1"

/db_xref="gi:12852333"

/db_xref="MGD:MGI:1921276"

/translation="MTMRSAFYKAAAPAGCNPPORLDYERAAALGPDESSGAEAHFLPHRKRLKEPPLASSOGSPSPASPGCGRGCLLPAGAAPQOESMKGSVPLPCPPATKQAGIGIEPVNAGSCSPSPRTQAVLPIDTGSTLVAAAEPTWADCKGSAAPAATAADDPAGPPPLPGPPPLPATATGLAASEGRMSIRKSPLGGGGSGASQAACAKIOILDLQDLLEOOQOLQAKETELSELSRDYLAIEMERMQLVKRDNKRNRKLKIQEPEEREAEALSEKIKLERQPLETSQALSFSGRSGKHKRRRPENTERNKNVYKRLAPEEFSEVKTTRKHSIKRECPGSISETYCKRELROEPFKRRSSVDTPRLSTPOKGFSTRHEKAFFSEMEDLYLETTEYLDRNHQPPSPPLPIARRSSPKKEEVNARCLMPSSVAGETSYLAVPWSMDHSVDFNPFDILENDVSFSRRHAKELEDERKKRMDIQRIREORIQLRLMYRKKGIOESEPEVTSEFPDDVESILTFLEFLVVAFGRLPLKPAQNFEFLPMIDERSRCRLIOKKHTPHTRCRK"

BASE COUNT

721 a 855 c 911 g 674 t

ORIGIN

Alignment Scores:

Pred. No.: 3161

Score: 3110.00 Length: 591

Percent Similarity: 97.56% Matches: 591

Best Local Similarity: 95.94% Conservative: 10

Query Match: 96.05% Mismatches: 13

ID: 11 Indels: 2

Gaps: 2

US-10-054-935-2 (1-614) x AK014463 (1-3161)

OY 1 Methrmetatgseralavalphelysalaalaaalaproialaglyglyasnproglu 20

Db 316 ATGCACCATGCAATCCGCGACTGTTCGAAGCGCGCCGCGCCCTGCGCGGGCAACCCGAG 375

OY 21 GlnargrleuaspyryguarAlaAlaAlaAlaenuglygyProgiuaspGluProglyVala 40

Db 376 CAGGAGCTGGACTACGAGCGGGGCTGGCGCTGGGGCGGGCCGAGAGCAAGTCCGGGGCG 435

OY 41 AlaAluaAlaHisPhaleuProArGHIsArqylsleuysGluPuProglyProProLeuala 60

Db 436 GCCGAGGCCATTTCTCCGCCCGGCATGTGAAGCTCAAGAGAGCGGGGGCCCCGCGCTGGCC 495

OY 61 SerSerGnuglyglySerProialaProSerProialaacylys--GlyglyLyglylaryg 79

[illegible]

Db	1576	GAGATGCAAGACTTTGGCGTACCTTTCCACCCACACAAATGTAATTGTGTGCTGGCACCAC	1635	
Qy	439	ProProProSerProLeuProLeuArgIuSerSerProIlySgluIuThrValAla	458	
Db	1636	CCTCCCCCATACCGTTTACATTCAGGGAATCCTCTCCAAAGAAAGAGACACTGTAGCA	16955	
Qy	459	ArgCysLeuMetProSerSerValAlaIgluIuThrSerValIleuAlaIvalProSerTrp	478	
Db	1696	AGGAGTCTGAGGCATCAACAGTGTGGACAGAGAAACCTCAGCTTGCGCTGTTCTTCTGG	17555	
Qy	479	ArgAspHisSerValGluProLeuAlaArgspProAsnProSerAspIleuLeuIuAsnLeu	498	
Db	1756	AGGGAGCCCTCAGTAGACCTCTCTAAGGAGCCCAATCCTTCAGACATTGTGGAGAACCTG	1815	
Qy	499	AspAspSerValPheSerIlyArgHisAlaIlySleuIuLeuAspGluIlyArgArgIys	518	
Db	1816	GATGACAGTGTATTTTCCAAAGAGCATGCCAAACTGGAGCTGACAGACAGAAAGAGAAAA	1875	
Qy	519	ArgTrpAspIleGlnArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMet	538	
Db	1876	CGATGGGATATTACAGAGGATCAGGAGAACAAAGAAATTTTACAGCGACTCAGCTCAGATG	19355	
Qy	539	TyrlsIlySlySlyGlyIleGlnIuSerIuProGluIuValIThrSerPhePheProGluPro	558	
Db	1936	TATATAAAGAAAGAAAGAAATTCAGAAATCTTAGCGCTGAGGTACTCCTATTTTCCCTGAGCA	19955	
Qy	559	AspAspValIgluSerLeuMetIleThrProPheLeuProValIAlaPheGlyIArgPro	578	
Db	1996	GATGATGTGTGAAAGTTTCTCTGATTTACCCCTTCTTGCTGTGTGACATTGGACGGCCA	20555	
Qy	579	LeuProIlySleuThrProGlnAsnPheGluLeuProTrpIleuAspGluIuArgSerArgCys	598	
Db	2056	TTGGCAAAATATTAGCTCCACAGAACCTTTGAATACCTGCTGGATGAGACGAACGGTTGC	21155	
Qy	599	ArgIleuGluIleGlnIlySlyGlnIThrProHisArgThrCysArgIys	614	
Db	2116	AGGTTGAGATCCAGAAAGACACACACCTCACCAGACCTGTAGAGAA	2163	
RESULT 2	AK015496	1304 bp	linear	HTC 19-JAN-2002
LOCUS	AK015496	Mus musculus adult male testis cDNA, RIKEN full-length enriched		
DEFINITION		library, clone:4930463F05:hypothetical protein, full insert		
ACCESSION	AK015496			
VERSION	AK015496.1	GI:12853861		
KEYWORDS	HTC; CAP trapper;			
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,			
	clone.lib:RIKEN full-length enriched mouse cDNA library			
	clone:4930463F05.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
JOURNAL	Carninci, P. and Hayashizaki, Y.			
MEDLINE	High-efficiency full-length cDNA cloning			
PUBMED	Meth. Enzymol. 303, 19-44 (1999)			
TITLE	99279253			
	10349636			
	2			
	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
	Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
	Normalization and subtraction of cap-trapper-selected cDNAs to			
	prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE	3			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,			
	Kono, H., Akiyama, J., Nishi, K., Kitanaka, T., Tashiro, H., Itoh, M.,			
	Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,			
	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,			
	Fujiwara, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watabiki, M.,			

QY 406 GYProSerThrHisProLysGluValAlaPheSerSerGluIleGluAspLeuProTyr 425
 DB 548 GGACCCACACCCACCCACAGAGAAAGCCCTTCTCAAGTGAATGAAGATTTGCCGATC 607
 QY 426 LeuSerThrThrGluMetTyrLeuGlySerArgTyrPheIleGlnProProProSerProLeuPro 445
 DB 608 CTTTCCACACAGAAATGTAATTTGTGTGGTGGACCAAGCCCTCCCATTCACCGTTACCA 667
 QY 446 LeuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSer 465
 DB 668 TTACGGGAATCTCTCCAAAGAGAGAGACTGTA----- 703
 QY 466 ValAlaGluThrSerValLeuAlaValProSerTyrArgAspHisSerValGluPro 485
 DB 704 -----GCATTCCTTTGGAGGACCACTGATAGAGCT 739
 QY 486 LeuArgAspProAsnProSerAspLeuLeuGluAsnLeuAspAspSerValPheSerLys 505
 DB 740 CTAAAGGACCAAAATCTTCAAGACATTTTGGAAAGCTGGATGACAGTATTTCCAAAG 799
 QY 506 ArgHisAlaLysLeuGluLeuAspGluLysArgTyrGlySerTyrPheIleGlnArgIle 525
 DB 800 AGCGATGCAAAAGCTGAGCTAGACAGAGAGAGAGAAACGATGGATATTCAGAGATC 859
 QY 526 ArgGluGlnArgIleLeuGluThrGluGlnLeuArgMetTyrLysLysGlyIleGln 545
 DB 860 AGGGAACAAGAAATTTTACAGCGACTGCAGCTCAGAAATGATATAAAGAGAAATTCAG 919
 QY 546 GluSerGluProGluValThrSerPhePheProGluProAspAspValGluSerLeuMet 565
 DB 920 GAATCTGAGCTGAGAGTATTCATTTTCCCTGAGCCAGATGATGTAAGTTGGCG 979
 QY 566 IleThrProPheLeuProValValAlaPheGlyArgProLeuProLysLeuThrProGln 585
 DB 980 ATTACCCCTCTCTGCTGTTGATGATTTGGACGCGCATTTGCCAAATATTCCTCCACAG 1039
 QY 586 AsnPheGluLeuProTyrPheLeuAspGluArgSerArgCysArgLeuGluIleGlnLys 605
 DB 1040 AACTTGAACACTACCTGTTGATGATGAGCAAGCCCTTGAGAGTTCAGAGTCCAGAGAG 1099
 QY 606 GlnThrProHisArgThrCysArgLys 614
 DB 1100 CACACACCTCACCGGACCTGTAGAGAG 1126
 RESULT 3
 B0014717/c 765 bp mRNA linear EST 26-MAR-2002
 LOCUS UI-H-ED1-ext-j-04-0-UI.s1 NCI_CGAP_ED1 Homo sapiens cDNA clone
 DEFINITION IMAGE:5833443 3', mRNA sequence.
 B0014717
 ACCESSION B0014717.1 GI:19739618
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 765)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Seq primer: M13 FORWARD
 Polya=Yes.
 FEATURES Location/Qualifiers

source 1..765
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5833443"
 /clone_lib="NCI_CGAP_ED1"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Left Pubic Bone; Vector: pMT3-Pac
 (Pharmacia) with a modified polylinker; Site:1: EcoR I;
 Site:2: Not I; NCI_CGAP_ED1 is a normalized cDNA library
 containing the following tissue(s): Chondrosarcoma cell
 line C55. The library was constructed according to Bonaldo
 & Lennon and Soares, Genome Research, 6:791-806, 1996.
 First strand cDNA synthesis was primed with an oligo-dT
 primer containing a Not I site. Double stranded cDNA was
 ligated to an EcoR I adaptor, digested with Not I, and
 cloned directionally into pMT3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is GCTCAAGCT.
 TAG_L1B=UI-H-ED1
 TAG_TISSUE="chondrosarcoma"
 TAG_SEQ="GCTCAAGCT"
 BASE COUNT 158 a 176 c 180 g 250 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,85e-71 Length: 765
 Score: 1324.00 Matches: 251
 Percent Similarity: 99.60% Conservative: 0
 Best Local Similarity: 40.89% Mismatches: 1
 Query Match: 14 Indels: 0
 Gaps: 0
 US-10-054-935-2 (1-614) x B0014717 (1-765)
 QY 291 ArgGluGluThrGluLeuSerGluLysIleLysLeuGluGlnProGluLeuSerGlu 310
 DB 763 AGAGAGGAAGACAGACTATCTGAGAAATTAACGTGAGTGCACCGAGCTTCCGAG 704
 QY 311 ThrSerGlnThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyHisLys 330
 DB 703 ACATCCAGACCTCTCCCTCCCAAGCCCTTCTCATTTGGGCGAGAGGAAAGACATATA 644
 QY 331 ArgLysSerProPheGlySerThrGluArgLysThrProValLysLeuAlaProGlu 350
 DB 643 AGGAATATCCCATTTTGAAGTACAGAAAGAAAGACTCGTTAAAGAAAGCTGCTCTGAA 584
 QY 351 PheSerLysValLysThrLysThrProLysHisSerProIleLysGluGluProCysGly 370
 DB 563 TTTTCAAAAGTCAAAACAAAACCTCTAAGCACTCTCATTTAAAGAGAAACCTGTGT 524
 QY 371 SerLeuSerGluThrValLysLysArgGluLeuArgSerGlnGluThrProGluLysPro 390
 DB 523 TCCTTATCTGAACACTGTTGTAAAGCTGATGAGAGACCAAGAAACCCCAAGAAAGCC 464
 QY 391 ArgSerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHis 410
 DB 463 CGGTCTTCAAGTGGACACCCACCAAGACTCTCCATCCCAAAAGGAGACCCAGACCAT 404
 QY 411 ProLysGluLysAlaPheSerSerGluIleGluAspLeuProTyrLeuSerThrThrGlu 430
 DB 403 CCCAAGGAGAAAGCCTTCTCAAGTAGAGTAAGAAATTTGGCTTACCTTCCACCAAGAA 344
 QY 431 MetTyrLeuCysArgTyrPheIleGlnProProProSerProLeuArgLysSerSer 450
 DB 343 ATGTAATTTGTGCTGTGGACACAGCTCCCATCATTACCTTACGGAATCTCTCT 284
 QY 451 ProLysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyLysThr 470
 DB 283 CCAAGAGAGAGAGAGACTGATGACAGAGTGTGATGATCATCAAGTGTTCAGAGAGAAACT 224

OY	471	SerValleuAlaValProSerTrrprkysphisservAlGluProLeuAraAsproksn	490
DB	223	TCAGCTTGCGCTGTTCCTCTTTGGGGACCACTCAGTAGAACCTCTTAAGGGACCANAAT	164
OY	491	ProSerAspleuleuGluaSnLeuAspaSserValPheSerTyARGHISAlaLysleu	510
DB	163	CCTTCAGACCTTTTGGAGAACCTGGATGCAGTGTGTTTTCGAACGGCATGCCANAAC TG	104
OY	511	GluLeuAspgLUlysaRgArqLyIsarGrtpasplIeglnatRglIeaRGluGlnaRgile	530
DB	103	GAGCTGATGTAGAAGAAAGAGAAAAGATATTATTCAGAGATCAGGAGCAAGAAAT	44
OY	531	LeuGlnaRgLeuGlnLeuAraMetyrtyrsLyslys 542	
DB	43	TTACAGCACCTGCAGCTCAGATGTATATAAAAAAAAA 8	
RESULT 4	BQ229745	870 bp mRNA linear EST 02-MAY-2000	
LOCUS	AGNCOURT_7510022 NIH_MGC_92 Homo sapiens CDNA clone IMAGE:6042882		
DEFINITION	5'-UTR mRNA sequence.		
ACCESSION	BQ229745		
VERSION	BQ229745.1 GI:20411145		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: cga@ds-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLML1382 row: p column: 19 High quality sequence stop: 596. Location/Qualifiers 1..870. /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image:6042882" /clone_id="NIH_MGC_92" /tissue_type="embryonal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."		
BASE COUNT	171 a 289 c 302 g 107 t	1 others	
ORIGIN			
Alignment Scores:			
Pred. NO.:	7.09e-70	Length:	870
Score:	1302.00	Matches:	260
Percent Similarity:	92.07%	Conservative:	7
Best local Similarity:	89.66%	Mismatches:	14
Query Match:	40.21%	Indels:	9
DB:	14	Gaps:	1
OY	US-10-054-935-2 (1-614) x BQ229745 (1-870)		
DB	13 AlapRoMaIagIyGIyaSnProGluGlnaRgLeuAspyrGIuargAlaAlalaLeucly 32		
	2 GCCCCTGCCGGCGGCAATCTCGACAGCGACTGGACTACGAGCGGGCTGCGCCTGGGC 61		

QY	33	GLYPRGGLUAPGILUPROGILYALALAGLUALAHISPHLEUPROAIRHISARGYLSEU	52
Db	62	GGGCCCCGAGAGACGACCTGGGGGGCCGAAAGCCACTTCTCCCGGACCCGTAAGCTC	121
QY	53	LYSGIUPROGILYPROPROLEUALASERSEINGILYGLYSERPROALAPROSERPROLA	72
Db	122	AAGAGAGCCGGGGCCCCCGCTGGCTCTCTCCAGGGCGGGAGGCCCGCTTCCCCGGCC	181
QY	73	GLYCSGLYGLYLYSGIYARGIYLEULEULEUPROALAGIYALALAPROGILYINGIN	92
Db	182	GGCTCGGGCGGCAAGGGCCGGGGCTGTATTACCTCCGGCGGGGGCGGCCCGGACGAG	241
QY	93	GIUGIUSERTPGIYGLYSERYVALPROLEUPROCYSPROPROPROALATHRYLSINLA	112
Db	242	GAAAGACCTGGGGGGGTTCGGTGGCTTGGCTTGGCCCCCGGCAACGAAGAACCC	301
QY	113	GLYIIEGLYGLYGLUPROALALALAGIYALAGIYCYSSERPROAIRPROLYSTYRGIN	132
Db	302	GGCATTGGGGGGGAGACCTCCCGCAGCCGGAGCCGGCTCAGGCCCGGCGCAAGTATCAG	361
QY	133	ALAYALLEUPROIIIEGLINTHRGLYSERLEUVALAIAALAYLSGLUPROTHRPROTP	152
Db	362	CGCGTGCGCCCATCAACAGCGGCTCTCTCGHGGCGGGCGCAAGAACCTTACGGCC	421
QY	153	ALAGIYAPPLYSGIYGLYALALASERPROALAIATHRALASERASPPROALAGIYPRO	172
Db	422	GCTGGGGCAAGGGGGGGGGCCCTCCCGCTGCAACCGCTCGGACCGGGGGACCC	481
QY	173	PROPROLEUPROLEUPROGILYPROPROPROLEUALAPROTHRALATHRALAGIYTHREU	192
Db	482	CCACACACTACCTGCCCCGGGGCCGACACCTCGGGCCACCGCCACCGCGCGGACCTG	541
QY	193	ALALASERGIUGIYAIRGTRPLYSSEMETARGLYSERPROLEUGIYGLYGLYGLY	212
Db	542	CGGGCCAGCGAGGGGAGATGGAAGAGTGTAGGAAGAGACCTCTCGGGGGGTGGCGGC	601
QY	213	SERGIYALASERSEINGILYALALACYSLEULYSGLIIELEULEULEULEULASPLEU	232
Db	602	TCGGGAGCTTCAGTCAGGCGCGCTCGCTCAACAGATCTTCTGCTCAATTGGACCTC	661
QY	233	IIEGLINGINGINGINGINLEUNGINALALYSGIULYSGIUIIEGLIULEULYSER	252
Db	662	ATCAACACGACGACGACACACTGACGGCCAAAGAAAGGAGATGAGAGCTGAATCA	721
QY	253	GLIARGASPHIREULEU-ALANRGLIEGLU-ARGMETGIUATARGMETGLINLEUVAL	271
Db	722	NAAAAGAAAAAGCGCTCTTCTGCTGGATTGAACGATGGAAGAGCGGGATGCACCGTG	781
QY	272	LYSLYSASPAANGIULYSIUGARHISLYSLEUPHEGINGIYTYR-----	286
Db	782	GTAAGAAGAGGATATACGAGAAAAAGAAAGAGGGCCACCAAGCTGTGTTTCAAGGGCCTA	841
QY	287	--GLUTHRGLIUGIUGIUGLU 293	
Db	842	TTAAAAACTTGAAAAAAAGAGAA 865	
RESULT 5			
BI665707		854 bp	mRNA
LOCUS			linear
DEFINITION	603287711.F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5321835 5'		EST 12-SEP-2001
ACCESSION	BI665707		
VERSION	BI665707.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		

QY	130	LysylarginylalaValleuProIleGlnThrglySerleuValAlaAlaIalysGluPro	149
Db	361	AAGATATAGCGGCTGCTGCCCATTTCAACAGGGCTCTCTGTGGGGGGCCAAAGAGCT	420
QY	150	ThrProTTPAlaGlyAspLysGlyGlyAlaAlaSerProAlaIarhAlaSerAspPro	169
Db	421	ACGGCTGTGGCTGGGACAAAGGGTGGGGGGCTCCCGACGTGCCACCGCTCGGACCG	480
QY	170	AlaGlyProProProProlleuProleuProGlyProProProlleuAlaProThAlaThAla	189
Db	481	GCGGGACCCACACCACTACCTGTGCCGGGGCGGACACCTCCGGGCCACACCACTGCC	540
QY	190	GlyThrleuAlaAlaSerGluGlyArgTrpLysSerMetArgLysSerProleuGlyGly	209
Db	541	GGAGACCTGGGGCGGCACTAGAGGACAGATGCAAGAGTATTAAGAAGAGACCCCTCTGGGGGT	600
QY	210	GlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeuYsglnIleLeuLeu	228
Db	601	GGCGCGGCGCTCGGAGCGCTCCAGTCAGGCGCGCTCAACACAGATCTTCTGCTG	657
RESULT 8			
LOCUS	B0178560	695 bp	EST 30-Apr-2002
DEFINITION	U1-M-EV0-bws-d-05-0-UI.r1 NIH-BMAP-EV0 Mus musculus CDNA clone		
ACCESSION	IMAGE:5701204 5', mRNA sequence.		
VERSION	B0178560		
KEYWORDS	B0178560.1 GI:20354052		
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 695)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	<p>Contact: Robert Strausberg, Ph.D. Email: cga@bts-remail.nih.gov Tissue Procurement: Dr. James Lin, University of Iowa CDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNML at: http://image.llnl.gov</p> <p>This clone was contributed by the Brain Molecular Anatomy Project (BMAP)</p> <p>Seq primer: PYX-5. Location/Qualifiers 1..695</p>		

BASE COUNT	98 a	264 c	244 g	88 t	1 others
ORIGIN					
Alignment Scores:					
Pred. No.:	6,11e-58			Length:	695
Score:	1104.00			Matches:	209
Percent Similarity:	97.25%			Conservative:	3
Best Local Similarity:	95.87%			Mismatches:	4
Query Match:	34.10%			Indels:	2
	14			Gaps:	2
US-10-054-935-2 (1-614) x BQ178560 (1-695)					
QY	1	MethrmetarqserAlaValPheLysAlaAlaAlaProAlaGlyGlyAsnProGlu	20		
Db	41	ATGACCAATGAAATCCCGCAGCTTTCAGAGCGCGCCCGGCTCTCGCGCGCAACCCGAG	100		
QY	21	GlnArgLeuAspTryGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAla	40		
Db	101	CAGGACACTGGACTACGAGCGGGGCTGGCGCTGGGCGGGCCGAGACGATCCGGGGCG	160		
QY	41	AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAla	60		
Db	161	GCCGAGGCCATTTCCTCCCGCGCATCGTAAGCTCAAGAGCGGGGGCCCGCTGGCC	220		
QY	61	SerSerGlnGlyGlySerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg	79		
Db	221	TCTTCCCGAGCGGGAGCGCCCTCGCCCTCCACCGCGCTCGCGCGCGCAAGGCGCGG	280		
QY	80	GlyLeuLeuLeuProAlaGlyAlaAlaAlaProGlyGlnGlnGluSerTryPglYlySer	99		
Db	281	GGCTGTGTACTCCAGCGGGGGGGCCCCCGGAGCAGAGAGAGAGCTGGGGGCTTCG	340		
QY	100	ValProLeuProCysProProProAlaAlaThrLysGlnAlaGlyIleGlyGlyLupProAla	119		
Db	341	GTGCCCTTGCCCTGTCCGCCGCCCGCAGCTACTAAACAGCGGCATCGCGGGAGCCACTC	400		
QY	120	AlaAlaGlyAlaGlyCysSerProArgProLysTryGlnAlaValaLeuProIleGlnThr	139		
Db	401	GCAGCGGCGCTGGGTGGAGCGCCCGGCCCAAGATACAGGGCGGTGCCATTACAGACG	460		
QY	140	GlySerLeu---ValAlaAlaAlaLysGluProThrProTrpAlaGlyAspLysGlyGly	158		
Db	461	GGCTCTATACGGTGGGGCGGGCCCAAGAGGCTTCGCGCTGGGGGAGCAAGGGTGGG	520		
QY	159	AlaAlaSerProAlaAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro	178		
Db	521	CGGGCTCCCCAGCTGCCACCGGCTCGGACCCGGCGGAGACCCCAACCACTACTCTGCCC	580		
QY	179	GlyProProProLeuAlaAlaProThrAlaAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArg	198		
Db	581	GGGCGACACACCCCTCGCGGCCACCGCCACTCTGGGANNCTGGGGGCGCAGTAGAGGCGAGA	640		
QY	199	TryPlySerMetArgLysSerProLeuGlyGlyGlyGlyGlySerGlyAlaSer	216		
Db	641	TGGAAGAGTATAAGAGAAAGAGCCCTCTCGGGGTGGCGCGGCTCGGAGGCTCC	694		
RESULT 9					
LOCUS	BQ257208	686 bp	linear	EST 06-MAY-2002	
DEFINITION	NISC_Koille07.q3 Baker mouse embryo e6.5 Mus musculus cDNA clone				
ACCESSION	BQ257208				
VERSION	BQ257208.1	GI:20457961			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 686)				
	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .				
	National Cancer Institute, Human Genome Anatomy Project (CGAP),				
	Tumor Gene Index				

ACCESSION B0187281
VERSION B0187281.1 GI:20362832
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 723)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msouares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 REVERSE.

FEATURES
source

1..723
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-a)-b-15-0-01"
/clone_lib="UI-E-EJ1"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site.1: Ecor I; Site.2: Not I; UI-E-EJ1 is a subcloned CDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, ACATCAGCA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT 231 a 171 c 170 g 148 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4.88e-56 Length: 723
Score: 1073.00 Matches: 208
Percent Similarity: 99.05% Conservative: 0
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 33.14% Indels: 1
DB: 14 Gaps: 0

US-10-054-935-2 (1-614) x B0187281 (1-723)

QY 249 Glnleuylserserglutarasprthrleuleualarqllleqluarqmetgluarqmet 268
|||||
Db 2 GACTGTAGACGAGAGAGACAGCTCTGCTGCGATTACACGTAT-GAAAGCGGATG 60
QY 269 GlnleuValylsaspasngluLysgluarqhisLysleupheglnlyTyrgluThr 288

|||||
Db 61 CAGCTGTAAGAAGGATTAACGAGAAAGAACGACAAAGCTGTTCCAGGGCTATGAACCT 120
QY 289 GlnluarqglugluThrprgileuSerGluylsleyleuGluylucysglnprgileu 308
|||||
Db 121 GAAGAGAGAGAGAGAGAGAGAGCTATCTGAAATAATTAACTGAGAGTCCAGCCGAGCTT 180
QY 309 SerGluThrserGlnThrleuProProlyserProPheSerGlyArgSerGlyGly 328
Db 181 TCGAGACATCCAGACTGCGCTCCCAACCCCTCTCATGTGGGGGAGATGAGAGGA 240
QY 329 HisLysArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeuAla 348
Db 241 CATAAAGAAATCCCATTTGGAGATACAGAAAGAAAGACTCTGTTAAAAAGCTGGCT 300
QY 349 ProgluPheSerLysValylsThrLysThrProLysHisSerProIleLysGluPro 368
Db 301 CCTGAATTTTCAAAAGTCAAAACAAACAAACTCTTACGACTCTCTATTAAAGAGAACCC 360
QY 369 CysGlySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGluThrProGlu 388
Db 361 TGTCGTCTTATCTGAACCTGTTTGAACGTGAATTGAGAGCCAAAGAACCCAGAA 420
QY 389 LysProArgSerSerValasprthrProProArgLysSerThrProGlnLysGlyProSer 408
Db 421 AAGCCCGGCTTCAGTGAGACACCCCAACCAACTCCACTCCCAAAAAGGAGACCCAGC 480
QY 409 ThrHisProLysGluLysAlaPheSerSerGluIleGluLysPheProTyrlsSerThr 428
Db 481 ACCCATCCCAAGAGAAAGCCCTTCTCAAGTAGATTAACATTTGGCTTTCACCC 540
QY 429 ThrGluMetTyrlsLeuCysArgTrpHisGlnProProProSerProLeuProLeuArgGlu 448
Db 541 ACGAATATGATTGTTGCTGCTTGGACACCACTCCCAACACACGTTACATTACGGGA 600
QY 449 SerSerProLysLysGluGluThrValAla 458
Db 601 TCCTCTCCANAGAGAGGAGACTGTAGCA 630

RESULT 13
LOCUS B1154331 811 bp mRNA linear EST 05-JUL-2001
DEFINITION 602905281.F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5034548 5',
B1154331
ACCESSION B1154331
VERSION B1154331.1 GI:14614332
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 811)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: L1AM1096 row: b column: 21
High quality sequence stop: 689.
Location/Qualifiers
1..811
/organism="Mus musculus"
/strain="129,C57BL/6J FVB/N"
/db_xref="taxon:10090"
/clone_image="5034548"
/clone_lib="NCI_CGAP_Mam3"

		Issue type="tumor, gross tissue"	/dev_stage="10 months"	/lab_host="DH10B"	/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
BASE COUNT	188 a 245 c 253 g 125 t				
ORIGIN					
Alignment Scores:					
Score: No.:	8,99e-56	Length:	811		
Percent Similarity:	1069.50	Matches:	237		
Best Local Similarity:	89.21%	Conservative:	11		
Query Match:	85.25%	Mismatches:	18		
	33.03%	Indels:	13		
	13	Gaps:	1		
US-10-054-935-2 (1-614) x B1154331 (1-811)					
OY	98	glySerAlaProLeuProCysProProProAlaThrLysGlnAlaGlyIleGlyGlu	117		
Dd	1	GGTTCGGTCCCTTGCCCTGTGCGCCCCACACTTAACAAGCGGCATCGGCGGAG	60		
OY	118	ProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProIle	137		
Dd	61	CCAGTCGACGCGCGCGTGGGTGACAGCCCGGCCCAAGTATCAAGCGGTCTCCATT	120		
OY	138	GlnThrGlySerLeu---ValAlaAlaAlaLysGluProThrProTTPAlaGlyAspLys	156		
Dd	121	CAGACGGGCTTATCGGTGGTGGCGCGCGCCAAAGACCTACGCCCTGGCGCGGAGAA	180		
OY	157	GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProLeuPro	176		
Dd	181	GGTGGGGGCTCCCGCCAGCGTCCAGCCGCTCGGAGCCCGGGGAGCCCGCCACTACT	240		
OY	177	LeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlu	196		
Dd	241	CTGCGCGGGCCACACCCCTGGCGCCACCGCCACTCTGGGACCTTGGCGCCAGTAC	300		
OY	197	GlyArgTyrPlysSerMetArgLysSerProLeuGlyGlyGlyGlyGlySerGlyAlaSer	216		
Dd	301	GGCAGATGGAGAATATAAGAGAAAGACCCCTCTCGGGGTGGCGCGCTCGGAGCCTC	360		
OY	217	SerGlnAlaAlaCysLeuLysGlnIleLeuLeuLeuGlnLeuAspLeuIleGluGlnGln	236		
Dd	361	AGTAGGCGCGCTCCCTCAACAACAATCTTCTGTGCATTTGGACTCTATCGAACACAG	420		
OY	237	GlnGlnGlnLeuGlnAlaLysGlnLysGluIleGluGlnLeuLysSerGlu-ArgAspThr	256		
Dd	421	CAGACGACGTTGCAGCGCCAGAGAAAGAGATTAAGAGAGCTGAAGTCCGACACAGATAC	480		
OY	256	TLeuLeuAlaArgIleGluArgMet-GluArgArgMetGlnLeuValLysLysAspAsnG	276		
Dd	481	GCCTCTTGCTCGGATTTGAACGTATAGGCAACGCGGATGACAGCTGATGAAGAGGATAACG	540		
OY	276	LulysGluArgHisLysLysLeuPheGlnGlyTyrGluThrGln-GluArgGluGluThrGlu	295		
Dd	541	AGAAAGGCAAGGCAACAAGCTGCTCCAGGGGCTATGAACTGACGACAGAGAGAAAGCCAG	600		
OY	296	LeuSerGluLys-LIeLysLeuGlnLysGlnPro-GluLeu-SerGluThrSerGlnThr	314		
Dd	601	TTTGCTGAGAAACATTCAATTGGAGCGCCAGCGGAGCTTGGACAGAGACTCCAGGCT	660		
OY	315	LeuPro-ProLysProPheSerCysGlyArg-SerGly-LysGlnHisLysArgLysSer	333		
Dd	661	CTGCGCTTCGGAAGGCTTACTCATGTGGCGGAGAGTGAACATGGGAGACCAAGGAGCAAG	720		
OY	334	ProPheGlySerThrGluArgLysThrProValLys--LysLeuAlaProGluPheSerL	353		
Dd	721	CCCATTTGGAAATACAGAAAGAAACATCTCTTTAAAGAAAGCTGGCTCTCAATTTCAA	780		

QY	353	ysValysThrlysthrProlyshsSerPro	363
Db	781	CAGTCACACACAC-AGTCTTAAGCAATCTCCA	811
RESULT 14			
LOCUS	BG421221	1087 bp	mRNA
DEFINITION	60245171221 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4590002 5',		
ACCESSION	NRNA sequence.		
VERSION	BG421221		
KEYWORDS	BG421221.1 GI:13327727		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 1087)		
COMMENT	NIH-MGC http://mhc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L1CMI324 row: h column: 03 High quality sequence stop: 583.		
FEATURES	location/Qualifiers		
source	1. 1087		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4590002"		
	/clone_lib="NIH_MGC_14"		
	/tissue_type="renal cell adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: kidney; Vector: pOT7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using 2AP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	293 a 309 c 345 g 140 t		
ORIGIN			
Alignment Scores:			
Score: 4.97e-55	Length: 1087		
Score: 1059.50	Matches: 240		
Percent Similarity: 81.13%	Conservative: 18		
Best Local Similarity: 75.47%	Mismatches: 44		
Query Match: 32.72%	Indels: 19		
DB: 12	Gaps: 3		
US-10-054-935-2 (1-614) x BG421221 (1-1087)			
QY	107	ProAlaThrlysglnAlaGlyTlEgLygLygLuProAlaAlaAlaGlyAlaGlyCysSer	126
Db	2	CCGGCGACCAAGCAAGCCGCGCATTCGGGGGAGCCTGCCGACCGGACCGCGCTGCACG	61
QY	127	ProArgProLysTYrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaA	146
Db	62	CCCCGGCCCAAGATACAGCGGTCGTCGCCATTCAGACGGCGCTCTCTGTCGGCGCGCC	121
QY	147	LysGluProThrProTrrProAlaGlyAspLysLysLygLyAlaAlaSerProAlaAlaThra	166
Db	122	AAAGGCGCTACGCCCTGGGCTGGGGACAAAGGTTGGGGGCGCTCCCGCGCTGCACGCG	181

QY 167 SerASPProAlaGlyProProProLeuProGlyProProProLeuAlaProThr 186
 |||||
 Db 182 TCGAGCCCGCGGAGCCCGCCACCACTCTCCGCGGCGCCACCCCTCGCGCCACC 241
 QY 187 AlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTrpLysSerMetArgLysSerPro 206
 |||||
 Db 242 GCCACCGCGGAGCCCTGGCGGCGAGGAGGAGATGGAAGATGAGGAAGAGACCCCT 301
 QY 207 LeuGlyGlyGlySerGlyAlaSerSerGluAlaAlaLysLeuLysGlnLeu 226
 |||||
 Db 302 CTCGGGGGTGGTGGCGGCTCGGAGGCTCCAGTCAGCGGCCCTCGCTCAAGAGATCCTT 361
 QY 227 LeuLeuGlnLeuAspLeuIleGluGlnGlnGlnGlnGlnGlnGlnGlnGln 246
 |||||
 Db 362 CTGCTCAATGGACCTCATCGAAGACAGACAGACAGCTGCGAGCCCAAGAGAGAG 421
 QY 247 IleGluGluLeuLysSerGluArgAspThrLeuAlaAlaGlyIleGluArgMet--GluA 266
 |||||
 Db 422 ATCGAGAGAGCTGAAGTCAGAGAGACGCTCTTGTCTCGATTAACCTATGCGCAAC 481
 QY 266 rGArgMetGlnLeuValLysLysAspAsnGlnLys--GluArgHisLysLeuPheGln-GI 285
 |||||
 Db 482 GCGGATGACAGCTGTAAGAAAGATACAGAGAACGAAAGACAGCAGCTTTTCAAGGG 541
 QY 285 yTrpGluThrGluGluArgGluGluThrGluLeuSerGluLysIleLysLeuGluGlu 305
 |||||
 Db 542 CATATGACACTGACGAGAGAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGCGCTTGA 601
 QY 305 nProGluLeuSerGluThrSerGlnThrLeuProProLysProPheSerCysGlyArgse 325
 |||||
 Db 602 GCG-GAGCTTCCGAGACATC-CAGACTCTGATGCCAAGCCCTTCTCATGTGGCGGAG 659
 QY 325 rGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrProValLys 345
 |||||
 Db 660 TGGGAGAGGAGCTTAACGGGAATCCAGAGAGAGAGTCAAGAAAGATCCGGGACAA 719
 QY 345 sLysLeuAlaProGluPheSerLys--ValLysThrLysThrProLysHisSerProIleL 365
 |||||
 Db 720 CAGCGCGGCGCCCAAGTTTACAGAGAGTGCAGACAAAGAGAGAGCTTAAGAGCTACTATCA 779
 QY 365 ys-GluGluProCysGlySerLeuSerGluThrValLys--LysArgGluLeuArgSer 383
 |||||
 Db 780 AGCGAGACACCATGTGGAATCCCTACTCGACACG-TGTGGAGACGATCACATGGCAGC 838
 QY 384 GlnGluThrProGluLysProArgSerSer-----ValAspThrProPro 398
 |||||
 Db 839 AACGAAACTCAAGAGAGACAGCGGCGCATACGAGAAACACACAGAGAGACTCCA 898
 QY 399 ArgLeuSerThrProGlnLysGlyProSerThrHisProLys 412
 |||||
 Db 899 TCGCTCTCT-----GGTGGCAGACCCGTCCAAG 928
 RESULT 15
 A1102093 615 bp mRNA linear EST 08-JAN-1999
 LOCUS EST111382 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
 DEFINITION RBR8Y80 3' end, mRNA sequence.
 ACCESSION A1102093
 VERSION A1102093.1 GI:4134057
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 615)
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Oct 6, 1998 this sequence version replaced gi:3706926.
 JOURNAL COMMENT
 Contact: Lee, NH

The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igr.org
 Seq primer: M13-21.
 Location/Qualifiers
 source
 1. 615
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RBR8Y80"
 /clone_lib="Normalized rat brain, Bento Soares"
 /note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 84 a 235 c 224 g 71 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.77e-55 Length: 615
 Score: 1055.50 Matches: 197
 Percent Similarity: 97.55% Conservative: 2
 Best Local Similarity: 96.57% Mismatches: 4
 Query Match: 32.60% Indels: 1
 DB: 9 Gaps: 1
 US-10-054-935-2 (1-614) x A1102093 (1-615)
 QY 13 AlaProAlaGlyLysAsnProGluGlnArgLysAspTrpGluArgAlaAlaLeuGly 32
 |||||
 Db 3 GCCCTGCGGCGGCGGCAACCCCGAGCAGCAGCTGACTGAGAGCGGCGCTCGGGC 62
 QY 33 GlyProGluAspLysProGluAlaAlaGlnAlaHisPheLeuProArgHisArgLysLeu 52
 |||||
 Db 63 GGGCCGAGAGACAGATCCGGGGCGGCGGAGAGCCACTCTCCCGCGCATGTAAGCTC 122
 QY 53 LysGlnProGlyProProLeuAlaSerSerGlnGlyLysSerProAlaProSerProAla 72
 |||||
 Db 123 AAGGAGCCCGGCGGCGGCGGCTGCTTACCAAGGAGGAGCCCGCGGCTCTCCAGCT 182
 QY 73 GlyCys--GlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaProGlyGln 91
 |||||
 Db 183 GGTGGGGGGCGGCGGAGGCGGCGGCTTTGTTACTCCGCGGCGGCGGCGGCGG 242
 QY 92 GlnGluLysSerTrpLysLysSerValProLeuProCysProProAlaThrLysGln 111
 |||||
 Db 243 CAGGAGAGAGAGCTGGGGCGGCTTGGCTTGCCTTGTCCGCGCGGCTACCAACAA 302
 QY 112 AlaGlyTLeuGlyLysLysProAlaAlaAlaGlyAlaGlyCysSerProArgProLys 131
 |||||
 Db 303 GCCGCGATCGCGGCGGAGCCAGTCGACGCGGCGGCTGCGACGCCCGGCGCAAGTAT 362
 QY 132 GlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaAlaLysGluProThrPro 151
 |||||
 Db 363 CAGCGGTGCTGCGCATTCAGACGGGCTCTCTGTGGCGGCGGCGCAAGAGCTACGCC 422
 QY 152 TrpAlaGlyAspLysGlyLysAlaAlaSerProAlaAlaThrAlaSerAspProAlaGly 171
 |||||
 Db 423 TGGGCTGGGGAGCAAGGCTGGGGCGGCTCCCGGAGCTGCGACCGCTCGGAGCGGGA 482
 QY 172 ProProProLeuProLeuProGlyProProProProAlaProThrAlaThrAlaGlyThr 191
 |||||
 Db 483 CCCCGACCACTACTTGGCCCGGCGGCGGCAACCCCTGCGGCCACCGCATCGCGGAGCC 542
 QY 192 LeuAlaAlaSerGluLysArgTrpLysSerMetArgLysSerProLeuGlyGlyGly 211
 |||||
 Db 543 CTGGCGGCGCAGTGAAGGAGATGAAGATATAGAAAGAGCCCTCTCGGGGTGGCGGC 602
 QY 212 GlySerGlyAla 215
 |||||
 Db 603 NGCTCGGAGGCC 614

Search completed: July 14, 2003, 22:10:00

Tue Jul 15 10:16:09 2003

us-10-054-935-2.rst

Page 15

Job time : 2357 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 5650 Seconds
(without alignments)
12532.154 Million cell updates/sec

Title: US-10-054-935-1

Perfect score: 4372

Sequence: 1 cagcctcgcagcccccgcac.....aaaaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1615406 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1950.4	44.6	3161	11	AK014463 Mus muscu
2	980.8	22.4	1304	11	AK015496 Mus muscu
3	897	20.5	1096	13	BM547318 BM547318 AGENCOURT
4	770.8	17.6	817	13	B1917709 B1917709 AGENCOURT
5	756.2	17.3	765	14	B0014717 B0014717 UI-H-ED1-
6	745.8	17.1	1032	13	BM475963 BM475963 AGENCOURT

Result No.	Score	Query Match	Length	DB ID	Description
7	744.6	17.0	891	14	BQ898842 BQ898842 AGENCOURT
8	737.4	16.9	931	14	BQ948768 BQ948768 AGENCOURT
9	725.6	16.6	870	14	BQ229745 BQ229745 AGENCOURT
10	714	16.3	715	14	BM981391 BM981391 UI-CF-FNI
11	711	16.3	729	14	B0004676 B0004676 UI-H-E10-
12	685	15.7	697	13	B1820795 B1820795 603034002
13	679.8	15.5	864	12	BG289135 BG289135 602387523
14	678.2	15.5	745	12	BE907276 BE907276 601500975
15	663.2	15.2	933	14	BQ876347 BQ876347 AGENCOURT
16	658.4	15.1	758	12	BE792962 BE792962 601585622
17	654.4	15.0	951	12	BG178033 BG178033 603227595
18	648.6	14.8	854	13	B1665707 B1665707 603287711
19	648.6	14.8	950	12	BG386583 BG386583 602454831
20	648.2	14.8	686	14	BM970469 BM970469 UI-CF-E01
21	638.4	14.6	640	12	BE896063 BE896063 601438862
22	637.4	14.6	686	13	B1914733 B1914733 603184105
23	625.8	14.3	1087	12	BG421221 BG421221 602451712
24	621	14.2	723	14	BQ187281 BQ187281 UI-E-E01-
25	619	14.2	627	14	BQ183599 BQ183599 UI-H-EUO-
26	618.8	14.2	686	14	BQ257208 BQ257208 NISC_K011
27	615.4	14.1	686	10	AM965141 AM965141 EST377214
28	613.4	14.0	891	14	BQ881177 BQ881177 AGENCOURT
29	613.2	14.0	889	12	BG506834 BG506834 601861447
30	605.2	13.8	710	10	AV713546 AV713546 AV713546
31	604.4	13.8	617	14	BM770559 BM770559 K-EST0054
32	604.4	13.8	695	14	BQ178560 BQ178560 UI-M-EV0-
33	597.6	13.7	637	10	BE693616 BE693616 MR4-BT035
34	597.6	13.7	647	12	BG537537 BG537537 602565862
35	597.6	13.7	698	13	B1905419 B1905419 603167481
36	595.8	13.6	631	10	BE693636 BE693636 MR4-BT035
37	595.6	13.6	705	10	BE277857 BE277857 601120091
38	595	13.6	629	10	AV722833 AV722833 AV722833
39	594.4	13.6	608	14	BQ636732 BQ636732 h013006_Y
40	591	13.5	677	12	BG180581 BG180581 602329333
41	590.6	13.5	662	12	BF284375 BF284375 EST448966
42	587.2	13.4	623	13	B1457952 B1457952 603198945
43	585.8	13.4	770	12	BG106391 BG106391 602289763
44	573.2	13.1	621	13	B1032095 B1032095 PM4-MT042
45	573	13.1	585	13	B1019744 B1019744 IL3-MT026

ALIGNMENTS

RESULT 1
LOCUS AK014463 3161 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:4121402D02:PRICHTEXTEN domain containing protein, full insert sequence.
ACCESSION AK014463.1 GI:12852332
VERSION AK014463.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 16 days embryo head cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20493774
PUBMED 11042159


```
|||||
Db 719 TCGTCCATTACAGCGGCTATCTGTTGGGGGGCCAAAGACTACGCCCTGGG 778
QY 535 CTGGGCAAGGTTGGGGGGCTCCCCCTGACACGGCTGGAGACCCGGGAGCCC 594
Db 779 CTGGGGCAAGGGTGGGGGGCTCCCGACCTGCGACGGCTGGAGCCGGGAGCCC 838
QY 595 CACCACTACCTTGCGGGGGCGGCGACCTCGGGGCCACCGGCGGGAGCCCTGG 654
Db 839 CACCACCTCTGCGGGGGCGGCGACCTCGGGGCCACCGGCGGGAGCCCTGG 898
QY 655 CGGCGACGAGGGCGAGATGGAAGATGGAAGAGAGCCCTCGGGGGTGGTGGCT 714
Db 899 CGGCGAGTGAAGGCGAGATGGAAGATGGAAGAGAGCCCTCGGGGGTGGGCGGCT 958
QY 715 CGGGAGCTCCAGTCAAGCGGCTGCTCAAAACAGATCTTCTGCTCAATTGGACCTCA 774
Db 959 CGGGAGCTCCAGTCAAGCGGCTGCTCAAAACAGATCTTCTGCTCAATTGGACCTCA 1018
QY 775 TCGAACAGCAGCAGCAGCAGCTGAGGCGAAGGAAAAGATGAGAGCTGAGTCA 834
Db 1019 TCGAACAGCAGCAGCAGCAGCTGAGGCGAAGGAAAAGATGAGAGCTGAGTCA 1078
QY 835 AGAGAGACAGCTCTCTGCTGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAG 894
Db 1079 AGAGAGATCGCTCTCTGCTGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAG 1138
QY 895 AGATATACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954
Db 1139 GGGATATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
QY 955 AAACAGAGCTATCTGAGAAATTAACAGAGTCCAGCGGAGCTTCCAGAGATCCC 1014
Db 1199 AAGAGAGTGTCTGAGAAATTAACAGAGTCCAGCGGAGAGAGAGAGAGAGAGAG 1258
QY 1015 AGACTGCTCCAGAGCCCTTCTCATGTTGGGGGAGAGAGAGAGAGAGAGAGAG 1074
Db 1259 AGGCTGCTCCAGAGCCCTTCTCATGTTGGGGGAGAGAGAGAGAGAGAGAGAG 1318
QY 1075 CCCCATTTGAGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1134
Db 1319 CCCCATTTGAGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378
QY 1135 AAGTCAAAACAAAACCTCTAGAGCTCTCTATTAAGAGAGAGAGAGAGAGAGAG 1194
Db 1379 AAGTCAAAACAAAACCTCTAGAGCTCTCTATTAAGAGAGAGAGAGAGAGAGAG 1438
QY 1195 CTGAACTGTTTGAACGTGAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
Db 1439 CAGAACTGTTTGAACGTGAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1498
QY 1255 CAGTGGACACCCCAAGAGCTCCACCTCCCAAAAGGAGAGAGAGAGAGAGAGAG 1314
Db 1499 CAGTGGATCCCAAGAGCTCCACCTCCCAAAAGGAGAGAGAGAGAGAGAGAGAG 1558
QY 1315 AGAAAGCTTCTCAAGTGAATGAAATTTGCCGTACTTCCACACAGAAATGTATT 1374
Db 1559 AGAAAGCTTCTCAAGTGAATGAAATTTGCCGTACTTCCACACAGAAATGTATT 1618
QY 1375 TGTGTGTTGGACACAGCTCCCATCACCCTTACATTAAGGGAATCTCTCAAGA 1434
Db 1619 TGTGTGTTGGACACAGCTCCCATCACCCTTACATTAAGGGAATCTCTCAAGA 1678
QY 1435 AGGAGGAGAGTGTAGCAAGGTTGATGACATCAAGTTTGCAGGAGAAATCACTCT 1494
Db 1679 AGGAGGAGAGTGTAGCAAGGTTGATGACATCAAGTTTGCAGGAGAAATCACTCT 1738
QY 1495 TGGCTGTCTCTTGTGAGGAGACACTCAGTGAAGCTCTAAGGAGAGAGAGAGAG 1554
Db 1739 TGGCTGTCTCTTGTGAGGAGACACTCAGTGAAGCTCTAAGGAGAGAGAGAGAG 1798
QY 1555 ACCTTTGGAGAACCTGATGACAGTGTGTTTGAAGGGAGAGAGAGAGAGAGAGAG 1614
|||||

Db 1799 ACATTTTGGAGAACCTGATGACAGTGTATTTTCAAAAGAGAGAGAGAGAGAG 1858
QY 1615 ATGAGAGAGAGAGAGAGAGAGATTTACAGAGAGATGAGAGAGAGAGAGAGAGAG 1674
Db 1859 ACAGAGAGAGAGAGAGAGAGATTTACAGAGAGATGAGAGAGAGAGAGAGAGAG 1918
QY 1675 GACTGAGCTCAGAGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
Db 1919 GACTGAGCTCAGAGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
QY 1735 CATTTTCCCTGAGCCAGATGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Db 1979 CATTTTCCCTGAGCCAGATGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
QY 1795 TAGCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Db 2039 TAGCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
QY 1855 ATGAGCTGAGCCGATGAGATGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Db 2099 ATGAGCTGAGCCGATGAGATGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
QY 1915 GGAATATGCTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1973
Db 2159 GGAATATGCTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2217
QY 1974 GAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2033
Db 2218 GAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2277
QY 2034 TACCTTTCACTACTTCTCTGA-----CTTGTGTTTCACTACTGATTTACAA----- 2082
Db 2278 TGCCTTCCATCCCTCTTATATATGATTTGTTTCTTCACTGATTTACAAACAA 2337
QY 2083 -----AACTCTTATTCGCTAATTGTGACTATGAGAGAGAGAGAGAGAGAG 2135
Db 2338 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
QY 2136 TCCCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2195
Db 2395 TCCATTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2450
QY 2196 CAGGGGTCTGTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2255
Db 2451 CAGGGGTCTGTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
QY 2256 ACATGTGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2315
Db 2507 -CGTGTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2565
QY 2316 CCCGTGTACACCCCTCTATAGTGTATGATTTTAAACCTTAATAATAACAAACA 2375
Db 2566 CCCGTGTACACCCCTCTATAGTGTATGATTTTAAACCTTAATAATAACAAACA 2604
QY 2376 ACCTCACCATGAGCTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435
Db 2605 ---TCACGGTGAAGCTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2656
QY 2436 TCTTCAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2495
Db 2657 TCTTCAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2698
QY 2496 TTTCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2555
Db 2699 TTTCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2717
QY 2556 TTTTCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2612
Db 2758 TTTTCAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817
QY 2613 GAAAGTGGAGATGATTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2671
Db 2818 AGAGATTGGAGATGATTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2877
```

QY	2672	TTGGGATGGGATCCAAACCTGTATTAATGTCCTGTATTCACATCACTCTTCAAGTATT	2731
Db	2678	TGGGATGGGATCCGAGAGCTGTAGAGATGCTCTG--TCATGCACTTTTCAAGATTT	2935
QY	2732	CCTTCATGCGGCTTCATCCCTTTAGACAGACTCTGTGGTGGGATAGAGACTTAGGAG	2791
Db	2936	CTTATGAGGGGCTTATCTATTTAGCCAACTTTTGGGTGGAGTGAAGATTTGGGGTG	2995
QY	2792	GG--TAGGGGAGAGTGTGGAATAGTGTCTTGTGGTGGGAAATGTCATCTTGG	2849
Db	2996	GGGGCGGGGGTGGAGATGAGAGATGATGCTTCCCTGCGGAGATGTCACACTTG	3055
QY	2850	AAACAAAGAGTGTACCTAATGAGCTTCCATTCACCTTGTATAATA--ATTGTATG	2908
Db	3056	AAACAAAGAGTGTACCTAATGAGCTT--TATTCGTTTGTGTAATAATGATTTATAC	3112
QY	2909	GTTACATCTGTGCTCTGCTCCCTCCCTCCCTTGGTT	2942
Db	3113	GAAACATCTGTGCTCCATCTCTGCTCCCTTGGTT	3146
RESULT 2			
AK015496			
LOCUS			
DEFINITION	AK015496	1304 bp	mRNA
			linear
			HTC 19-JAN-2002
			library, clone:4930463F05:hypothetical protein, full insert
ACCESSION	AK015496		
VERSION	AK015496.1	GI:12853861	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4930463F05.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
PMID	99279253		
PMED	10349636		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
PMID	20499374		
PMED	11042159		
REFERENCE			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasaiyagi, K., Fujiwaka, S., Inoue, K., Iogawa, Y., Iizawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
PMID	20530913		
PMED	11076861		
REFERENCE			
AUTHORS	Kawai, J., Shitagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M.,		

Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bul, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashimo, J., Mazzarelli, J., Mombeerts, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.			
Functional annotation of a full-length mouse cDNA collection			
Nature 409 (6821), 685-690 (2001)			
PMID 12185660			
PMED 11217851			
5 (pages 1 to 1304)			
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bul, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Iizawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shitagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toy, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
Direct Submission			
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome.res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9216, Fax: 81-45-503-9216)			
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.			
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to 15' GAGAGAGAGAGATCCAGGATGCTTTT			


```

|||||
Db 181 TAGAGACTTAGGAGGAGTAGGGGAGAGTGGAAATAGTCTCTTGGCTGGCAAA 240
OY 2837 TGTCTACATCTTGAACAACAGATGACTATAGCTTCCATCTACTTGTAAAA 2896
Db 241 TGTCTACATCTTGAACAACAGATGACTATAGCTTCCATCTACTTGTAAAA 300
OY 2897 TAATTTGATGTGTACATCTTGGTCCCTCCCGTGGTTGTAAATATCAGANA 2956
Db 301 TAATTTGATGTGTACATCTTGGTCCCTCCCGTGGTTGTAAATATCAGANA 360
OY 2957 GCACGCCAGGCGCACTTGTCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 3016
Db 361 GCACGCCAGGCGCACTTGTCTCAGTGAAGATCCCTATTAATCTGGAAGAAAT 420
OY 3017 AGAGCCAGACCTCGGTCTCAATATATAGAAATGCTTCTTGTAGCTCAGAGCTA 3076
Db 421 AGAGCCAGACCTCGGTCTCAATATATAGAAATGCTTCTTGTAGCTCAGAGCTA 480
OY 3077 TTGTGTGAACAACAGTAGGGGTCTAATCTCCAGAGGAGTAGGGGCTTTATCTTAAAGA 3136
Db 481 TTGTGTGAACAACAGTAGGGGTCTAATCTCCAGAGGAGTAGGGGCTTTATCTTAAAGA 540
OY 3137 GAATATGTCCCAATATATAGCACTTTTGAAGAGAAAGCAAGGTATGTAGGGTGTG 3196
Db 541 GAATATGTCCCAATATATAGCACTTTTGAAGAGAAAGCAAGGTATGTAGGGTGTG 600
OY 3197 GCTGGCCCATCAGTGTAGCAGCAAGAGAAATGAGATACATTTGGGAGAGAAAGAAA 3256
Db 601 GCTGGCCCATCAGTGTAGCAGCAAGAGAAATGAGATACATTTGGGAGAGAAAGAAA 660
OY 3257 GTTCTCTCAGGGGCGCTCCACACTGCTAAAGTTTGTGAGATGTATCTGCTCTCTG 3316
Db 661 GTTCTCTCAGGGGCGCTCCACACTGCTAAAG-TTTTGTGAGATG-TGATCTGTCTCTCTG 718
OY 3317 ATTTCAGTTTAAAGAAATATTTGCGACACATGTATTTGTGATGTCTGCTG 3376
Db 719 ATTTCAGTTTAAAGAAATATTTGCGACACATGTATTTGTGATGTCTGCTG 778
OY 3377 CTCCTATTTCTCCCTTTTGTGCTGTGTGTGTG 3410
Db 779 TCTTATTTCTCTTGTGTGTGTGTGTGTG 812

RESULT 5
LOCUS B0014717/c 765 bp mRNA linear EST 26-MAR-2002
DEFINITION UT-H-ED1-axt-j-04-0-UI-s1 NCI_CGAP_ED1 Homo sapiens cDNA clone
IMAGE:5833443 3', mRNA sequence.
ACCESSION B0014717
VERSION B0014717.1 GI:19739618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLA-ies.
FEATURES
source 1..765
Location/Qualifiers

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5833443"
/clone_lib="NCI_CGAP_ED1"
/rnuce_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/Note="Organ: Left Public Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site:1: EcoR I;
Site:2: Not I; NCI_CGAP_ED1 is a normalized cDNA library
containing the following tissue(s): Chondrosarcoma cell
line C55. The library was constructed according to Bonaldo
Lennon and Soares, Genome Research, 6:791-806, 1996.
First strand cDNA synthesis was primed with an oligo-dT
primer containing a Not I site. Double stranded cDNA was
ligated to an EcoR I adaptor, digested with Not I, and
cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)18 tail. The
sequence tag for this library is GCTCAAGGCT.
TAG_LIB=UT-H-ED1
TAG_RTSUB=Chondrosarcoma
TAG_SEQ=CGTCAAGCT"
BASE COUNT 158 a 176 c 180 g 250 t 1 others
ORIGIN
Query Match 17 38; Score 756.2; DB 14; Length 765;
Best Local Similarity 99.58%; Pred. No. 5,7e-92;
Matches 758; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

946 AGAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGAGCCGAGGAGCTTCCG 1005
Db 765 AGAGAGAGAAACAGAGCTATCTGAGAAATTAACCTGAGAGCCGAGGAGCTTCCG 706
OY 1006 AGACATCCCAAGCTGTGCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGGACATA 1065
Db 705 AGACATCCCAAGCTGTGCTCCCAAGCCCTTCTCATGTGGGCGAGTGAAGGACATA 646
OY 1066 AAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACCTCTTAAAGAGCTGCTCTG 1125
Db 645 AAAGAAATCCCATTTTGAAGTACAGAAAGAAAGACCTCTTAAAGAGCTGCTCTG 586
OY 1126 AATTTTCAAAAGTCAAAACAAACCTCTAGACCTCTCTTAAAGAGAGAAACCTGTG 1185
Db 585 AATTTTCAAAAGTCAAAACAAACCTCTAGACCTCTCTTAAAGAGAGAAACCTGTG 526
OY 1186 GTTCTCTTCTGAAAGCTGTTGTAAAGCTGAATGAGAGAGCCAGAGAAAGC 1245
Db 525 GTTCTCTTCTGAAAGCTGTTGTAAAGCTGAATGAGAGAGCCAGAGAAAGC 466
OY 1246 CCCGCTCTCAAGTGAACACCCACCAAGACTCTCCACTGCCCAAAAGGAGCCAGACCC 1305
Db 465 CCCGCTCTCAAGTGAACACCCACCAAGACTCTCCACTGCCCAAAAGGAGCCAGACCC 406
OY 1306 ATCCAGAGAGAAAGCTCTGCAAGTGAATGAAGATTTGGCCGATACCTTCCACACAG 1365
Db 405 ATCCAGAGAGAAAGCTCTGCAAGTGAATGAAGATTTGGCCGATACCTTCCACACAG 346
OY 1366 AATGTATTTGTGTGTGGCAACAGCTCCCAATCAACCTTACCATTTAGGGAGATCT 1425
Db 345 AATGTATTTGTGTGTGGCAACAGCTCCCAATCAACCTTACCATTTAGGGAGATCT 286
OY 1426 CTCGAAGAGAGAGAGACTGTAGCAAGGTGTCTATGCCATCAAGTGTGAGAGAAA 1485
Db 285 CTCGAAGAGAGAGAGAGACTGTAGCAAGGTGTCTATGCCATCAAGTGTGAGAGAAA 226
OY 1486 CTTCAAGTGTGCTGTCTCTTCTTGGAGGAGCACTCAGTAGAGGCTCTTAAGGAGCCAA 1545
Db 225 CTTCAAGTGTGCTGTCTCTTCTTGGAGGAGCACTCAGTAGAGGCTCTTAAGGAGCCAA 166
OY 1546 ATCTTCAGACCTTTTGGAGAACCTGATGACAGTGTGTTTGAAGGAGGATGCAAAAC 1605
Db 166 ATCTTCAGACCTTTTGGAGAACCTGATGACAGTGTGTTTGAAGGAGGATGCAAAAC 111

```

Db 165 ATCCTTAGACCTTTTGAGACACCTGGATGCACATGTTGTTTCGACCGCATGCACAAAC 106

QY 1606 TGGAGCTGGATGATAAGAGAAAGAAAGATGGATATTCACAGAGATCGGGACAAAGAA 1665

105 TGGAGCTGGATGATAAGAGAAAGAAAGATGGATATTCACAGAGATCGGGACAAAGAA 46

Db 1666 TTTTACAGCGACTGCACCTCAGATGTATATAAAAGAAAGAA 1707

QY 45 TTTTACAGCGACTGCACCTCAGATGTATATAAAAGAAAGAA 4

RESULT 6	1032 bp	mrna	linear	EST 05-FEB-2000
BM475963				
LOCUS				
DEFINITION	AGGECOUNT 6480858	NIH_MGC_67	Homo sapiens	CDNA IMAGE:5579934
ACCESSION	BM475963			
VERSION	5', mRNA sequence.			
KEYWORDS	BM475963.1	GI:18525014		
SOURCE	EST.			
	human.			

REFERENCE AUTHORS TITLE JOURNAL COMMENT
<p>1 (bases 1 to 1032) NIH-MGC http://imgc.ncbi.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.</p>

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LRAM12337 row: g column: 07
 High quality sequence stp: 653.

FEATURES	Location/Qualifiers
source	1. .1032

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5579934"
/clone_id="NH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPO6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."

```

Query Match	17.1%	Score 745.8	DB 13	Length 1032
Best Local Similarity	94.4%	Pred. No. 1.2e-90		
Matches 852; Conservative	0	Mismatches 37	Indels 14	Gaps 7

QY 1495 TGGCTGTTCCCTTCTTGGAGGACCACTCAGTAGAGCCCTTAAGGGACCCAATCCTTCAG 1554

[illegible][illegible]

QY 1615 ATGAGAGAGAAGGAAGATGGATATTTCAGAGATCAGGACAACAAGATTTCACGC 1674

Db 156 ATGAGAGAGAGAAAGATGGGATATTCAGAGGATCAGGGAACAAGATTTCAGC 215

107.5 GACATCCAGCTCAGATATGATAAAAAGAAAGGAATTCAGGAATCTGAGCCTGAGGTTACCT 1/3

OY	1735	CATTTCGCCGAGCAGATATGTGTGAAATTTGATGATTAACCCCTCTTGGCTGTG	1794
Db	276	CATTTCGCCGAGCAGATATGTGTGAAATTTGATGATTAACCCCTCTTGGCTGTG	335
OY	1795	TAGCATTTGGAGCAGCATTACCAAAATTAACCTCACAGAAATTTTGAGCTACCTGTGTG	1854
Db	336	TAGCATTTGGAGCAGCATTACCAAAATTAACCTCACAGAAATTTTGAGCTACCTGTGTG	395
OY	1855	ATGAGCGTAGCGCGATGCGATTGGAGATCCGAAAGAAACCAACCCCTACGGGACGTGA	1914
Db	396	ATGAGCGTAGCGCGATGCGATTGGAGATCCGAAAGAAACCAACCCCTACGGGACGTGA	455
OY	1915	GGAATATAGCTGTGCTGGCAGAAACCCCTGTCTTCAGATAGTTGTAGCATGCAATCCGAG	1974
Db	456	GGAATATAGCTGTGCTGGCAGAAACCCCTGTCTTCAGATAGTTGTAGCATGCAATCCGAG	515
OY	1975	AGTGGGAGAGACCTGTATATGTGACCTTTGTCTCACAATATGTATATCACTGGCTGATAT	2034
Db	516	AGTGGGAGAGACCTGTATATGTGACCTTTGTCTCACAATATGTATATCACTGGCTGATAT	575
OY	2035	ACCCTTCATCTCCCTGCACTTGTGTTTTCATCTACTCGATTTCACAAAACCTGTTTCAT	2094
Db	576	ACCCTTCATCTCCCTGCACTTGTGTTTTCATCTACTCGATTTCACAAAACCTGTTTCAT	635
OY	2095	TCGGCTAATTTGTGACTATGTGAGGGTGTATGGATTTCTTTT--CCCTTTTGGGAAATG	2153
Db	636	TCGGCTAATTTGTGACTATGTGAGGGTGTATGGATTTCTTTTCCCTTTTGGGGAAGG	695
OY	2154	GGCTCTCAAGTAAAGCATATAGAAAGGAGATTCAGAAATTTCA--GGGGCTGTGTTTCAT	2212
Db	696	GGCTCTCAAGTAAAGCATATAGAAAGGAGATTCAGAAATTTCAAGGGGGTCTGTTTTCAT	755
OY	2213	ACA--TTTGCTATGTAAAGGGGTAAAGGGGCTCTCTTCATTAGACATGTGG--AAGATG	2269
Db	756	ACATTTTGGCTATGTAAAGGGGTAAAGGGGCTCTCTCTTCATTAGACATGTGGGAAGATG	815
OY	2270	AAGAGAGCCCTT--CCTTTAGAGCTGT--GCCGCAAGGACCTCTCTCAACCCCTGGATACCC	2327
Db	816	AAGAGAGCCCTTCCCTTTAAAGCTGTGGCCCTGCATGGAACCTCTCTCTCCCGGGGTACA	875
OY	2328	CTCCTTATAGTGGG-----TATAGTGAATTTTAAACCTAAATTAACCAACACCTC	2380
Db	876	CCCTCCCTTATAGGGGGGTATGTGATTTTAAACCTTAAATTAACCAACACACCTC	935
OY	2381	ACC 2383	
Db	936	CCC 938	

RESULT 7	LOCUS	DEFINITION
B0898842	B0898842	891 bp mRNA linear EST 16-AUG-2002
	AGENCOURT_8121821	Lupsk1.dorsal.root.ganglion Homo sapiens CDNA clone IMAGE:6181536 5' mRNA sequence.

VERSION BQ898842.1 GI:22290856

SOURCE ORGANISM

REFERENCE	1 (bases 1 to 891)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

Db	61	GTGATGCTGTGGCTATGGCTTTTCATTTGTAACCTCATTCTGTAAGAGAGATGGGCTCTC	120
Qy	3466	TATATAGGGAACCTGCTGTAACTTCATTTCGACGCAAGGATGTAGAGAGAAATAGACTTAA	3523
Db	121	TATATAGGGAACCTGCTGTAACTTCATTTCGACGCAAGGATGTAGAGAGAAATAGACTTAA	180
Qy	3526	TTCCACTATAGGGGCTCTCATCTCACACCTTAAGAGAGAAATTTCTAGAAAACTGGGCCAG	3585
Db	181	TTCCACTATAGGGGCTCTCATCTCACACCTTAAGAGAGAAATTTCTAGAAAACTGGGCCAG	240
Qy	386	ATTTTCTTTGTTCTCCATCATTTTAATGTGGCAGGCTGTCTCAGTTTCTTACTCTTACCT	364.5
Db	241	ATTTTCTTTGTTCTCCATCATTTTAATGTGGCAGGCTGTCTCAGTTTCTTACTCTTACCT	300
Qy	3646	ATGTGATTTTCTTCGTACGCTGTCCAAAAAGAAAAAGACCACATCATGTCTCTTCTAC	3705
Db	301	ATGTGATTTTCTTCGTACGCTGTCCAAAAAGAAAAAGACCACATCATGTCTCTTCTAC	360
Qy	3706	TTTGTCTTTTGATCCCTCAGTTTCTTCTGATTTTTCAGCATGTGTGGGTTTCTTATTTTG	3765
Db	361	TTTGTCTTTTGATCCCTCAGTTTCTTCTGATTTTTCAGCATGTGTGGGTTTCTTATTTTG	420
Qy	3766	GGTATGATTTGCAAAATTTAAACCAATGTGTTTGCCCTACCCAGGGGACCTCCAGATT	3823
Db	421	GGTATGATTTGCAAAATTTAAACCAATGTGTTTGCCCTACCCAGGGGACCTCCAGATT	480
Qy	3826	CTGACTTGAATAGACAGTGAAGAAGATCCAGCAGGCTCATCTGCGCAGATTTAAGTAGAT	3885
Db	481	CTGACTTGAATAGACAGTGAAGAAGATCCAGCAGGCTCATCTGCGCAGATTTAAGTAGAT	540
Qy	3886	TCTATTTTCTTGTGTTCTCCCTCTGCTGAGACCTCTTATTTATTTGTCCTCTTCTAG	394.5
Db	541	TCTATTTTCTTGTGTTCTCCCTCTGCTGAGACCTCTTATTTATTTGTCCTCTTCTAG	600
Qy	3946	GTTAATTTCTCTTTGATTTGACTTTGTGTGAGAGAGAGGTGTGACAGTATTAGCAAGT	4005
Db	601	GTTAATTTCTCTTTGATTTGACTTTGTGTGAGAGAGAGGTGTGACAGTATTAGCAAGT	660
Qy	4006	TTCAAAGTGCAGAAATTTACAGTGTGTAGAGTGTGGGGG--AAATTAATCTTAA--TTTTTC	4062
Db	661	TTCAAAGTGCAGAAATTTACAGTGTGTAGAGTGTGGGGGAAATTAATCTTAAATTTTTTC	720
Qy	4063	CTTACATGGGG--ATPACACACG--GTGAATTCATCTTCAACG--AAGGCCCTGCACTGCTC	4119
Db	721	CTTACATGGGNAATACACACTGGTGAATTCATCTTCACTGAAGAGGCCCTGCACTGCTC	780
Qy	4120	CTAAACATA---GTTGTGTTGTTTCTTTTCTTAACAA--GTTTAAGCTAGTGTAAATATA	4173
Db	781	TTTAAACCTAAAGTTGGGGTGTGTTTCTTCTTAACAAAGTTTAAAGTAAAGGTTAATATA	840
Qy	4174	TTTAAAAAAA---TTGCTGTGCTGTAC--TTACGCTTGTGTTTATGCCCCATT	4223
Db	841	TTTAAAAAAAATTTGGCTGTGCTGTGCACTTCAAGCTTTGTTTATGCCCCCAATT	895
RESULT 9			
LOCUS	BQ229745	870 bp	mrna
DEFINITION	AGNCNCURF_7510022 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042882	linear	EST 02-MAY-2002
ACCESSION	BQ229745		
VERSION	BQ229745.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eumalypota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 870)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

FEATURES	source
Base Count	171 a 289 c 302 g 107 t 1 others
Query Match	16.6%; Score 725.6; DB 14; Length 870;
Best Local Similarity	99.0%; Pred. No. 6.4e-88;
Matches	761; Conservative 0; Mismatches 5; Indels 3; Gaps 3
Tissue	Email: c9aphs-rf@mail.nih.gov
Tissue Procurement	ATCC
CDNA Library Preparation	Life Technologies, Inc.
CDNA Library Arrayed by	The I.M.A.G.E. Consortium (ILNL)
CDNA Sequencing by	Agencourt Bioscience Corporation
Clone distribution	MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov	
Plate:	LLNL13282 row: p column: 19
High quality sequence stop:	556.
Location/Qualifiers	1. 870
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_image:6042882"
	/clone_id="NIH_MGC_92"
	/tissue_type="embryonal carcinoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: testis; Vector: pCMV-Sport6; site_1: NotI; site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

QY 773 CATGACAGCAGCAGCAGCTGACAGCCAGGAAAAAGATCGAGCTGAGTC 832
 |||||||
 Db 661 CATGACAGCAGCAGCAGCAGCTGACAGCCAGGAAAAAGATCGAGCTGAGTC 720
 |||||||
 QY 833 AG-AGAGAGACACGCTCC-TTGCTCGATTG-AACGTATGGAAGCGCG 878
 |||||||
 Db 721 ANAAGAGAAAAAGCCTCTTGTCTGCGATTGAAACGTATGGAAGCGCG 769
 |||||||
 RESULT 10
 BM981391/c 715 bp mRNA linear EST 21-MAR-2002
 LOCUS UI-CF-EN1-adh-f-11-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
 DEFINITION UI-CF-EN1-adh-f-11-0-UI 3', mRNA sequence.
 ACCESSION BM981391
 VERSION BM981391.1 GI:19603832
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 715)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul.mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 FORWARD
 POLYA=yes.
 FEATURES
 source Location/Qualifiers
 1. 715
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-adh-f-11-0-UI"
 /clone_11b="UI-CF-EN1"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-CF-EN1 is a normalized cDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pRTT3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CTGCTCAGCT.
 TAG_L1B=UI-CF-EN1
 TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 6hr to LPS 24h
 TAG_SEQ=CTGCTCAGCT"
 263 a 132 c 123 g 196 t 1 others
 BASE COUNT
 ORIGIN

Query Match 16.3%; Score 714; DB 14; Length 715;
 Best Local Similarity 99.9%; Pred. No. 2.6e-86;
 Matches 714; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3644 CTATGATATATTTCTGTAAGTGTCCAAAAAAGAACCAATCACTGCTCTG 3703
 |||||||
 Db 715 CTATGATATATTTCTGTAAGTGTCCAAAAAAGAACCAATCACTGCTCTG 656
 |||||||
 QY 3704 ACTTTGTTCTTGAATCCCTCAGTTCTTCTGATTTACAGATGTCGGTCTTAATT 3763
 |||||||
 Db 655 ACTTGTCTTGAATCCCTCAGTTCTTCTGATTTACAGATGTCGGTCTTAATT 596
 |||||||
 QY 3764 TGGGTATGATTAACAATTAACCATGTTGTTGTCCTACCAAGGAGCTCCAGT 3823
 |||||||
 Db 595 TGGGTATGATTAACAATTAACCATGTTGTTGTCCTACCAAGGAGCTCCAGT 536
 |||||||
 QY 3824 TTCTGACTTGAAGTGAAGTGAAGAAATCCAGAGTGTCTATCTGCGCAGATTAAGTAG 3883
 |||||||
 Db 535 TTCTGACTTGAAGTGAAGTGAAGAAATCCAGAGTGTCTATCTGCGCAGATTAAGTAG 476
 |||||||
 QY 3884 ATTCATATTTCTGTTGTTCTCCCTCTCCCTGAGACCTTTATTTATTTGTCCTCTCT 3943
 |||||||
 Db 475 ATTCATATTTCTGTTGTTCTCCCTCTCCCTGAGACCTTTATTTATTTGTCCTCTCT 416
 |||||||
 QY 3944 AGTTAATTTCTCTTGTATTTGACTTTGTGAGAAGGAGTTGACAGTAGATAGCAAA 4003
 |||||||
 Db 415 AGTTAATTTCTCTTGTATTTGACTTTGTGAGAAGGAGTTGACAGTAGATAGCAAA 356
 |||||||
 QY 4004 GTTCCAAAGTGAAGTGAAGTGAAGTGTGAGAGTGTGAGAGTGTGAGTGTGAGTGTG 4063
 |||||||
 Db 355 GTTCCAAAGTGAAGTGAAGTGAAGTGTGAGAGTGTGAGAGTGTGAGTGTGAGTGTG 296
 |||||||
 QY 4064 CTACATGGGATACAACTGTCGATTCATCTTCACTGAAGGCCCTGCACTTCTCTAA 4123
 |||||||
 Db 295 CTACATGGGATACAACTGTCGATTCATCTTCACTGAAGGCCCTGCACTTCTCTAA 236
 |||||||
 QY 4124 AACATAGTGTGTTGTTTCTTAAAGAGTTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4183
 |||||||
 Db 235 AACATAGTGTGTTTCTTAAAGAGTTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 176
 |||||||
 QY 4184 TTGCTGTCTGCTACCTTCACTGCTGTTTATGCCCATTTGATTTGTTGCTGTGTGT 4243
 |||||||
 Db 175 TTGCTGTCTGCTACCTTCACTGCTGTTTATGCCCATTTGATTTGTTGCTGTGTGT 116
 |||||||
 QY 4244 AATTCATACCTTTGATACCACTTCTGATGCTAAATTTGTTGTTGTTGTTGTTGTTG 4303
 |||||||
 Db 115 AATTCATACCTTTGATACCACTTCTGATGCTAAATTTGTTGTTGTTGTTGTTGTTG 56
 |||||||
 QY 4304 TAAAGAGTCAATTTGTAATTAACATTTGTCGCTGTAAAAAAGAAAAA 4358
 |||||||
 Db 55 TAAAGAGTCAATTTGTAATTAACATTTGTCGCTGTAAAAAAGAAAAA 1
 |||||||
 RESULT 11
 B0004676/c 729 bp mRNA linear EST 26-MAR-2002
 LOCUS UI-H-E10-ayp-e-01-0-UI.s1 NCI-CGAP_E10 Homo sapiens cDNA clone
 DEFINITION UI-H-E10-ayp-e-01-0-UI.s1 NCI-CGAP_E10 Homo sapiens cDNA clone
 IMAGE:5841768 3', mRNA sequence.
 ACCESSION B0004676
 VERSION B0004676.1 GI:19729576
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 729)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 The following repetitive elements were found in this cDNA
 sequence: 1-37, >POLY_A\$Simple_repeat (matched complement)
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES

source

Location/Qualifiers
 1. 729
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5841768"
 /clone_id="NCI_CGAP_E10"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
 with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
 NCI_CGAP_E10 is a cDNA library containing the following
 tissue(s): Chondrosarcoma. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is ACACCTGCAC.
 TAG_L1B=UI-H-E10
 TAG_TISSUE=chondrosarcoma
 TAG_SEQ=ACACCTGCAC"

BASE COUNT 248 a 141 c 137 g 203 t
 ORIGIN

Query Match 16.3%; Score 711; DB 14; Length 729;
 Best Local Similarity 100.0%; Pred. No. 6.4e-86;
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3476 CCGCGTGAACCTTCATTCGACGAGGATGTAGAGAAATAGACTTAATTCACATGAG 3555
 |||||||
 729 CCGCGTGAACCTTCATTCGACGAGGATGTAGAGAAATAGACTTAATTCACATGAG 670
 |||||||
 3536 GGCCTCATCTCACACTTAAGGAGGAGATTTAGAAAAAGGGGACATTTCTTTG 3595
 |||||||
 669 GGCCTCATCTCACACTTAAGGAGGAGATTTAGAAAAAGGGGACATTTCTTTG 610
 |||||||
 3596 TTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTTCTTCTACTTACTATGATAT 3655
 |||||||
 609 TTCTCCATCATTTTAATGTGGCAGGCTGTTCAGTTCTTCTACTTACTATGATAT 550
 |||||||
 3656 TCTTCGTACAGTGTCCAAAAAGAAAGCCCAATGAGTGTCTTGAATTTGTTT 3715
 |||||||
 549 TCTTCGTACAGTGTCCAAAAAGAAAGCCCAATGAGTGTCTTGAATTTGTTT 490
 |||||||
 3716 GATCCCGCATCTTCTTTGATTTACAGATGTGCGGCTCCTAATTTGGGATAGTT 3775
 |||||||
 489 GATCCCGCATCTTCTTTGATTTACAGATGTGCGGCTCCTAATTTGGGATAGTT 430
 |||||||
 3776 AGCAAAATTAACCATGTGTGTGTGGCCCTACCCAGGAGCTCCCACTTCTGACTGAA 3835
 |||||||
 429 AGCAAAATTAACCATGTGTGTGTGGCCCTACCCAGGAGCTCCCACTTCTGACTGAA 370
 |||||||
 3836 GTGAGCTGAGAAAGATCCAGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTTCT 3895
 |||||||
 369 GTGAGCTGAGAAAGATCCAGAGGTGCTATCTGGCCAGATTAAAGTAGATTCTTCT 310
 |||||||
 3896 TTGTTTCCTCTCCCTGAGAGCTCTTATTTATGTGCTCCCTCTTCTAGGTTAATTC 3955
 |||||||
 309 TTGTTTCCTCTCCCTGAGAGCTCTTATTTATGTGCTCCCTCTTCTAGGTTAATTC 250
 |||||||

QY 3956 CTTTGATTTGACCTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAGTTCCAGTGCA 4015
 |||||||
 DB 249 CTTTGATTTGACCTTTGTTGAGAGAGGTTGACAGTAGATTAGCAAGTTCCAGTGCA 190
 |||||||
 QY 4016 AATTTACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGATA 4075
 |||||||
 DB 189 AATTTACAGTGTGTAGAGTGTGGGGGAAATTAAGTCTTATTTTCCCTACATGGATA 130
 |||||||
 QY 4076 CAACACTGTAATTCATTTCACTGAGAGCCCTGCAGTTCCTTAAACATAGTTGT 4135
 |||||||
 DB 129 CAACACTGTAATTCATTTCACTGAGAGCCCTGCAGTTCCTTAAACATAGTTGT 70
 |||||||
 QY 4136 TGTGTTTCTTTAACCAAGTTTAAAGCTGTTAATAATTAATAAATTTG 4186
 |||||||
 DB 69 TGTGTTTCTTTAACCAAGTTTAAAGCTGTTAATAATTAATAAATTTG 19
 |||||||

RESULT 12
 BI820795/c 697 bp mRNA linear EST 04-OCT-2001
 LOCUS 603034002E1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5175097 5',
 DEFINITION mRNA sequence.
 ACCESSION BI820795
 VERSION BI820795.1 GI:15932345
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M1436 row: c column: 02
 High quality sequence stop: 693.

FEATURES

source

Location/Qualifiers
 1. 697
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5175097"
 /clone_id="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 6 male brains, age range 23-27; 1
 male lung, age 27; and 1 male testis, age 69. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.8 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 021. Note: this is a NIH-MGC Library."

BASE COUNT 210 a 167 c 118 g 202 t
 ORIGIN

Query Match 15.7%; Score 685; DB 13; Length 697;
 Best Local Similarity 99.9%; Pred. No. 1.9e-82;
 Matches 696; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2103 TTGAGATTATGAGGATGATGATTTCTTTCCCTTTTGGGAATGGGCTGCAC 2162
 |||||||
 DB 697 TTGAGATTATGAGGATGATGATTTCTTTCCCTTTTGGGAATGGGCTGCAC 638
 |||||||

FEATURES	LOCATION/Qualifiers	EST 21-FEB-2001
2163	GCTAAGCTATAGGATGGCAGATTCACAGAGTTTCAAGGGCTGTCTTATACATTTCCCT	2222
2163		
637	GCTTAAGCTATAGGATGGCAGATTCACAGAGTTTCAAGGGCTGTCTTATACATTTCCCT	578
2223	ATGTTAAAGGGGGTAAAAAGGGCTCTCTTCATTAGACATAGTGGAMATGAAGCAGCCCTTC	2282
577	ATGTTAAAGGGGGTAAAAAGGGCTCTCTTCATTAGACATAGTGGAMATGAAGCAGCCCTTC	518
2283	CTTTAGAGCTGTGCTTCGATGCGACTCTTCTCACCCGTGATACACCTTCCTTATAGTGGGT	2342
517	CTTTAGAGCTGTGCTTCGATGCGACTCTTCTCACCCGTGATACACCTTCCTTATAGTGGGT	458
2343	ATAGTGAATTTTAAACCCCTA-AAATTAACAAACAAACCTCACCATGAGCTTTAGAACCGA	2401
457	ATAGTGAATTTTAAACCCCTA-AAATTAACAAACAAACCTCACCATGAGCTTTAGAACCGA	398
2402	AGAGGAATGACAACTGAAAGCGATGAGCAAGCACCCTTCACAGAGTAAAAAGACATCGG	2461
397	AGAGGAATGACAACTGAAAGCGATGAGCAAGCACCCTTCACAGAGTAAAAAGACATCGG	338
2462	AGAGTGTGTGATACAGTCTGGAAGAATGTTGTCATTTGGAACATTCCTGATGATCA	2521
337	AGAGTGTGTGATACAGTCTGGAAGAATGTTGTCATTTGGAACATTCCTGATGATCA	278
2522	GTCATGTGGGAAGGATGTTTGGCTGTGATATTTTTCAGTTAATGGATMAACATTTCT	2581
277	GTCATGTGGGAAGGATGTTTGGCTGTGATATTTTTCAGTTAATGGATMAACATTTCT	218
2582	TTACTGCTCAAAAACCCAAATCTTTGGAAAAAGAGTGGATGTATGTTTCAGAACAA	2641
217	TTACTGCTCAAAAACCCAAATCTTTGGAAAAAGAGTGGATGTATGTTTCAGAACAA	158
2642	GTTTCACGCTGTAAACAAAGACACTTATTTTGGATGCGATGCGCAAAACCTGTATTAAT	2701
157	GTTTCACGCTGTAAACAAAGACACTTATTTTGGATGCGATGCGCAAAACCTGTATTAAT	98
2702	GTCCTTGTATCACATCACTTCTCAAGTATTTCTTCATTTGGGCTTCATCTTTAGCAGAA	2761
97	GTCCTTGTATCACATCACTTCTCAAGTATTTCTTCATTTGGGCTTCATCTTTAGCAGAA	38
2762	CTCTTGGTGGGATGAGACTTAAGGAGGGGTAGGG	2798
37	CTCTTGGTGGGATGAGACTTAAGGAGGGGTAGGG	1
RESULT 13		
Bg289135	864 bp	linear
LOCUS	602387523P1 NIH_MGC_93 Homo sapiens	cdna clone IMAGE:4516420 5',
DEFINITION	mRNA sequence.	
ACCESSION	Bg289135	
VERSION	Bg289135.1	GI:13044673
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
TITLE	1 (bases 1 to 864)	
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgabs-remail.nih.gov	
	Tissue Procurement: ATCC	
	cdna Library Preparation: Life Technologies, Inc.	
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	
	DNA Sequencing by: Incyte Genomics, Inc.	
	Clone distribution: MGC clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LNLN at:	
	http://image.lhnl.gov	
	plate: LHAM10407	row: n column: 05
	High quality sequence stop: 634.	
	location/Qualifiers	

Source	1. 864	/organism="Homo sapiens"	/db_xref="taxon:9606"	/clone="IMAGE:4516420"	/clone_1id="NIH_MGC_93"	/issue="transitional cell papilloma, cell line"	/lab_host="DH10B (phage-resistant)"	/note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo dt primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
BASE COUNT	236 a	171 c	199 g	258 t				
ORIGIN								
Query Match	15.5%	Score 679.8;	DB 12;	Length 864;				
Best Local Similarity	92.4%;	Pred. No. 8.4e-82;						
Matches 759;	Conservative 0;	Mismatches 57;	Indels 5;	Gaps 4;				
1972	GAGAGTGGCAGACGCTGTATATGACGCTTTGCCACATATGATATCAGCTGGCAT	2031	GAGAGTGGCAGACGCTGTATATGACGCTTTGCCACATATGATATCAGCTGGCAT	60				
2032	AATACCCCTTCATCTCTCTGACTTGTATTCTTCATTCACGATTTTCAAAAACCTCTT	2091	AATACCCCTTCATCTCTCTGACTTGTATTCTTCATTCACGATTTTCAAAAACCTCTT	120				
2092	CATTGCGCTAATTTGTAGTTATGGAGGGTATTTGGGATTTCTTTCCCTTTTGGGAAA	2151	CATTGCGCTAATTTGTAGTTATGGAGGGTATTTGGGATTTCTTTCCCTTTTGGGAAA	180				
2152	TGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAAAGTTTCAGGGGCTGTCTCTA	2211	TGGGCTCTCAAGCTAAAGCTATAGATGGCAGATTCAGAAAGTTTCAGGGGCTGTCTCTA	240				
2212	TACATTTTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCATTAGACATGTGGAAATGAA	2271	TACATTTTGCTATGTTAAAGGGGTAAAGGGCTCTCTTCATTAGACATGTGGAAATGAA	300				
2272	GCAGCGCCCTTCTTAAAGCTGCTCGTCGTCAGATGGACGACCTCTTCACCCGTTACACCTCC	2331	GCAGCGCCCTTCTTAAAGCTGCTCGTCGTCAGATGGACGACCTCTTCACCCGTTACACCTCC	360				
2332	TTAATAGTGGTATAGTATTTTAAACCTTAATTAACCAACAACCTCACCATAGACTT	2391	TTAATAGTGGTATAGTATTTTAAACCTTAATTAACCAACAACCTCACCATAGACTT	420				
2392	TAGAGCAAGAAGGAATGACAACTGMAAGCATATAGCAAGCCATCTTCACAGGTGAA	2451	TAGAGCAAGAAGGAATGACAACTGMAAGCATATAGCAAGCCATCTTCACAGGTGAA	480				
2452	AAGACATCGAGAGTTGGTAGATTAACGTCTGAAAAGATAGTGTTCATTGTAACATATT	2511	AAGACATCGAGAGTTGGTAGATTAACGTCTGAAAAGATAGTGTTCATTGTAACATATT	540				
2512	CTGTGATRCAGTCATGTGGGAAGGAGTTTGGCTGTGATTAATTTTTCAGTTAATGAT	2571	CTGTGATRCAGTCATGTGGGAAGGAGTTTGGCTGTGATTAATTTTTCAGTTAATGAT	599				
2572	AACAAATTTCTTACGCTC-AAAAACCAAAATCTTTGGAAAAAGAAAGGGGAGTGT--	2638	AACAAATTTCTTACGCTC-AAAAACCAAAATCTTTGGAAAAAGAAAGGGGAGTGT--	659				
2639	AGTTTCAGAACAGTTACAGCTGTAAACAAAGACCTAGTATTTGGAGTGCATGCGCAA	2698	AGTTTCAGAACAGTTACAGCTGTAAACAAAGACCTAGTATTTGGAGTGCATGCGCAA	718				
2699	GGTTTTCAGAACAGCTCACAGCTGT- AACCAAGACCTAGTATTTGGAGTGCATGCGCAA	2748	GGTTTTCAGAACAGCTCACAGCTGT- AACCAAGACCTAGTATTTGGAGTGCATGCGCAA	778				
2749	CCTTTTAGCAGAACTCTGTGCTGGTGGGATAGAGACTTAGGG	2789	CCTTTTAGCAGAACTCTGTGCTGGTGGGATAGAGACTTAGGG					

Db 779 TCTCTAGGCAAAATCCTGGGGGCGGAGCAAACTTAGAG 819

RESULT 14
LOCUS BE907276 745 bp mRNA linear EST 20-OCT-2000
DEFINITION 601500975F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902610 5', mRNA sequence.

ACCESSION BE907276
VERSION BE907276.1 GI:10400670
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 745)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LLM9705 row: f column: 19
High quality sequence stop: 691.
Location/Qualifiers
1. 745
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 194 a 127 c 187 g 237 t

ORIGIN

Query Match 15.5%; Score 678.2; DB 12; Length 745;
Best Local Similarity 96.8%; Pred. No. 1.5e-81;
Matches 692; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2780 AGACTTAGGAGGGGAGAGAGTGTGAATAGGTGCTTCCTTGGCTGCAATGT 2839
|||||
Db 1 AGACTTAGGAGGGGAGAGAGTGTGAATAGGTGCTTCCTTGGCTGCAATGT 60
|||||

QY 2840 CTACATCTTGAACAACAGATGATGAGCTTCCTTGCATTCCTTGAATAA 2899
|||||
Db 61 CTACATCTTGAACAACAGATGATGAGCTTCCTTGCATTCCTTGAATAA 120
|||||

QY 2900 TTGTATGTGTACCATCTTGCTCTCCCTCCCGTTTGTAAATATCAGATAGCA 2959
|||||
Db 121 TTGTATGTGTACCATCTTGCTCTCCCTCCCGTTTGTAAATATCAGATAGCA 180
|||||

QY 2960 CTCCAGGCGCATTTGGTCTCACTGTAGATCCCTTATTAATGAAAGAAATAGA 3019
|||||
Db 181 CTCCAGGCGCATTTGGTCTCACTGTAGATCCCTTATTAATGAAAGAAATAGA 240
|||||

QY 3020 GCCAAGACCTGCTGCAATATATAGAAATGCCCTTCTTATAGTCTTACAGATATG 3079
|||||
Db 241 GCCAAGACCTGCTGCTCAATATATAGAAATGCCCTTCTTATAGTCTTACAGATATG 300
|||||

QY 3080 TGTGAACAAGTAGGGGTCTAATCTCTGAAGTAGGGGCTTTATCTTAAAGAGAA 3139
|||||
Db 301 TGTGAACAAGTAGGGGTCTAATCTCTGAAGTAGGGGCTTTATCTTAAAGAGAA 360
|||||

QY 3140 TATGTCCCGCATTTATTAGCATTTTAGAGAGAACCAAGATATGAGGTGTGGCT 3199
|||||
Db 361 TATGTCCCGCATTTATTAGCATTTTAGAGAGAACCAAGATATGAGGTGTGGCT 420
|||||

QY 3200 GGGCCATCATGTGAGCAGCAGAGAGATGGATACCATTTGTGGAGAGAGAAAAAGTT 3259
|||||
Db 421 GGGCCATCATGTGAGCAGCAGAGAGATGGATACCATTTGTGGAGAGAGAAAAAGTT 480
|||||

QY 3260 CCTCAGGGGCGCTCCACATGCTTAAGTTTTTTGTGAGATGTGATCTGCTCTGGATT 3319
|||||
Db 481 CCTCAGGGGCGCTCCACATGCTTAAGTTTTTTGTGAGATGTGATCTGCTCTGGATT 540
|||||

QY 3320 TGACCTTTAAGGAATATTCTGCAGACATGATATTCTTGGATGATGCTGCTC 3379
|||||
Db 541 TGACCTTTAAGGAATATTCTGCAGACATGATATTCTTGGATGATGCTGCTC 600
|||||

QY 3380 TTATTTCTCTTTTGT 3439
|||||
Db 601 TTATTTCTCTTTTGT 660
|||||

QY 3440 CCATCTGCTTAGAGAGATGGGCTCTCTATAGGAGAACCTGCTTAACCTTATTG 3494
|||||
Db 661 CCATCTGCTTAGAGAGATGGGCTCTCTATAGGAGAACCTGCTTAACCTTATTG 715
|||||

RESULT 15
BO876347 933 bp mRNA linear EST 16-AUG-2002
LOCUS BO876347
DEFINITION AGENCOURT 8585335 Lupeki_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6192304 5', mRNA sequence.

ACCESSION BO876347
VERSION BO876347.1 GI:22268353
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 933)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupeki
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LLM13594 row: b column: 17
High quality sequence stop: 508.
Location/Qualifiers
1. 933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="Lupeki_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6 (Life Technologies); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors:
5'-TCGACCCAGCGCTCCG-3' and
5'-GACTAGTCTAGATCGCGAGCGGCCCTT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupeki, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

BASE COUNT 185 a 261 c 190 g 296 t 1 others

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 19:15:06 ; Search time 370 Seconds
(without alignments)
3737.100 Million cell updates/sec

Title: US-10-054-935-2
3338
Perfect score: 1 MTMRSAVFKAAPAGNPPE.....RSRCRLFIQKQPHRTCRK 614
Sequence:

Scoring table:
BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-x1h
-MODEL=framed_p2n_model -DB=us10054935/runat_14072003_115025_24619/app.query.fasta.1.775
-Q=/cgr2_1/USPFO.spool/US10054935/runat_14072003_115025_24619/app.query.fasta.1.775
-DB=us10054935.spool/US10054935/runat_14072003_115025_24619/app.query.fasta.1.775
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcr -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=us10054935.ecgn.1.1.263 &runat_14072003_115025_24619 -NCPU=6 -ICPU=3
-NO_MMAR -LARGESOURCE -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: N.Genesecg.101002: *
2: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1980.DAT: *
3: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1981.DAT: *
4: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1982.DAT: *
5: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1983.DAT: *
6: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1985.DAT: *
7: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1986.DAT: *
8: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1987.DAT: *
9: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1988.DAT: *
10: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1989.DAT: *
11: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1990.DAT: *
12: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1991.DAT: *
13: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1992.DAT: *
14: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1993.DAT: *
15: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1994.DAT: *
16: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1995.DAT: *
17: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1996.DAT: *
18: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1997.DAT: *
19: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1998.DAT: *
20: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA1999.DAT: *
21: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA2000.DAT: *
22: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA2001A.DAT: *
23: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA2001B.DAT: *
24: /SIDS2/gcgdata/genesecg/genesecg-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1448.5	44.7	1529	21 AAF18193	Lung cancer associ
2	1351.5	41.7	2887	21 AAF6874	Human ORF2429
3	1055.5	32.6	615	24 ABK62818	Rat sequence diffe
4	680.5	21.0	2887	21 AAC76874	Human ORF2429
5	246	7.6	7713	22 AAK51958	Human polynucleoti
6	245	7.6	5373	24 AAD30567	Human kinase poly
7	238	7.4	9551	20 AAZ22301	cDNA encoding a hu
8	236	7.3	7033	21 AAZ36989	Human polynucleoti
9	235.5	7.3	4037	22 AAK59242	Human polynucleoti
10	231	7.1	3608	24 AAD33243	Human secreted pro
11	228.5	7.1	3169	22 AAI58822	Human polynucleoti
12	228.5	7.1	3203	22 AAI60608	Human polynucleoti
13	225	6.9	2108	24 ABL67774	Oesophagus cancer
14	223.5	6.9	3117	21 AAZ38858	Human Jurkat cell
15	223.5	6.9	3414	23 AAS78091	DNA encoding novel
16	222.5	6.9	8459	22 AAC89557	Human histone deac
17	222.5	6.9	117213	19 AAV62176	HSV-2 strain SB5 C
18	221.5	6.8	1833	23 AAS67931	DNA encoding novel
19	221.5	6.8	7516	22 AAK51987	Human polynucleoti
20	220.5	6.8	8460	24 ABK87718	Human cDNA encodin
21	220	6.8	2438	24 ABQ54916	Human ovarian anti
22	220	6.8	6432	23 AAS65903	DNA encoding novel
23	220	6.8	13555	22 AAS34526	Human DNA for a no
24	220	6.8	13555	22 AAL04593	Human reproductive
25	220	6.8	13555	23 ABL97516	Human testicular a
26	220	6.8	154746	24 AAD25519	Human herpesvirus
27	220	6.7	154746	24 AAD25519	Human herpesvirus
28	218	6.7	6457	24 ABN96892	Gene #3390 used to
29	217.5	6.7	2614	23 ABL28430	Drosophila melanog
30	217.5	6.7	4954	23 ABL05187	Drosophila melanog
31	217.5	6.7	9516	23 ABL05186	Drosophila melanog
32	217	6.7	3464	23 ABL30199	Drosophila melanog
33	217	6.7	3959	24 AB199206	Mouse ischaemic co
34	217	6.7	5524	23 ABL30198	Drosophila melanog
35	217	6.7	36778	21 AAZ87318	S. venezuelae pik
36	217	6.7	37948	21 AAZ87318	S. venezuelae pik
37	217	6.7	38506	21 AAA75633	Nucleotide sequenc
38	217	6.7	38506	21 AAZ56001	Recombinant cosmid
39	216.5	6.7	3720	23 AAS74064	DNA encoding novel
40	216	6.7	3147	22 AAS03399	Gene #3351 used to
41	215.5	6.7	3147	22 AAS03399	Thermus DNA encodi
42	215	6.6	14835	24 AAS94858	Human DNA sequence
43	214.5	6.6	7746	21 AAA07836	Human homologue of
44	214.5	6.6	4411529	22 AAI99682	Mycobacterium tube
45	213.5	6.6	2725	24 ABK84458	Human cDNA differe

ALIGNMENTS

RESULT 1
AAFI8193
ID AAF18193 standard: DNA; 1529 BP.

AC AAF18193;
14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 212.

KW Human; lung cancer associated protein; neuroprotective; cytosolic;

KW cardioactive; immunomodulatory; muscular active; vulnerable;

KW gastrointestinal; nephrotoxic; anti-infective; gynecological;

KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

KW proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

XX

PN W0200055180-A2.
 XX 21-SEP-2000.
 PD
 PE 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 PI Ruben SM:
 XX WPI: 2000-587514/55.
 DR P-PSDB: AAB58317.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 XX Claim 1; Page 676-677; 1425pp; English.
 PS
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytoskeletal, cardiocactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotropic; antifibrotic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 XX Sequence 1529 BP; 460 A; 344 C; 379 G; 344 T; 2 other;
 SQ
 Alignment Scores:
 Pred. No.: 6,45e-43 Length: 1529
 Score: 1448.50 Matches: 284
 Percent Similarity: 95.64% Conservative: 1
 Best Local Similarity: 95.30% Mismatches: 6
 Query Match: 44.73% Indels: 7
 DB: 21 Gaps: 2
 US-10-054-935-2 (1-614) x AAF18193 (1-1529)
 QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTrpLys 200
 DB 3 CCACCCCTCGGCCGCCACCGCCGCGGACCTCGCGCGCCACGAGGAGATGGAAG 62
 QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAla 220
 DB 63 AGTATGAGAAAGACCTCTCTGGGGGTGGTGGCGGCTCGGAGCTCCACAGTCAAGCGGCC 122
 QY 221 CysLeuLysGlnLeuLeuLeuGlnLeuAlaPheLeuLeuGlnGlnGlnGlnGlnLeu 240
 DB 123 TGCCTCAAAACAGATCCTTCTGCTGCAATTGCACTTCATCGAACGACGACACACAGCTG 182
 QY 241 GlnAlaLysGluLysGluLysGluLysGluLysSerGluLysArgPheLeuAlaArg 260
 DB 183 CAGGCCAAGGAAAGGAGATCGAGAGACTCAAGTCAAGAGACAGCGCTCTTCTGCTGG 242
 QY 261 IleGluArgMetGluArgMetGlnLeuValLysLysAspAsnGluLysGluArgHis 280
 DB 243 ATTGAAGTATGGAAGGCGGATGCACTGTGTAAGAAGATATACGAGAAAGAAAGCGCAC 302

QY 281 LysLeuPheGlnGlyTrpGluThrGluLysArgGluLysGluLeuSerGluLysIle 300
 DB 303 AAGCTGTTTCAGGGCTATGAACTGAGAGAGAGAGAAACAGAGCTATCTGAGAAATT 362
 QY 301 LysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrLeuProPheProPhe 320
 DB 363 AAACGTGAGTGCACCGCCGAGACCTTCGAGACATCCACAGACTGCTCCCAACCCCTTC 422
 QY 321 SerCysGluArgSerGlyLysGlyLysArgLysArgLysSerProPheLysSerThrGluArg 340
 DB 423 TCATGTGGCGCGAGATGGAAGGACATGAAAGAAATCCCATTTGGAATACGAAAGA 482
 QY 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys 360
 DB 483 AAGACTCCTGTTAAAGAGCTGCTCGAATTTCAAAAGCAAAACAAACCTCTAAG 542
 QY 361 HisSerProIleLysGluGluProCysGlySerLeuSerGluThrValLysArgGlu 380
 DB 543 CACTCTCTTAAAGAGGAAACCTGTGGTCTTATCTGAACGTGTTGTAACGTGAA 602
 QY 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeu 400
 DB 603 TTGAGAGCCAGAAACCCAGAAAGCCCGGCTTCAGTGGACACCCCAAGACTC 662
 QY 401 SerThrProGlnLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIle 420
 DB 663 TCCACTCCCAAAAGGAGACCCAGACCCATCCCAAGAGAAAGCCTCTCAAGAGATGA 722
 QY 421 GluAspLeuProThrLysSerThrThrGluMetThrLeuLysArgGluArgPheGlnProPro 440
 DB 723 GAACATTTGGCGTACCTTTCCACCCAGAAATGATTGTGCTTGCCACCCACCCCTCC 782
 QY 441 ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCys 460
 DB 783 CCATCACCAGTATACCATTCAGGGAAATCTCTCCAAAGAGAGAGAGACTGTAGCA----- 836
 QY 461 LeuMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
 DB 837 -----ACTAAGCATAGAGAAACACTTGTCTT---ATACCTTAGTGG 875
 RESULT 2
 AAC76874
 ID AAC76874 standard; cDNA; 2887 BP.
 XX
 AC AAC76874;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORF2429 polynucleotide sequence SEQ ID NO:4857.
 XX
 KW Human; open reading frame; ORF; detection; cytoskeletal; hepatotropic;
 KW vulnary; antiproliferative; antiparkinsonian; nocotropic; neuroprotective;
 KW anticonvulsant; osteoprotic; antidiarrheal; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX

PR 31-MAR-1999: 99US-0127607.
 PR 02-APR-1999: 99US-0127636.
 PR 05-APR-1999: 99US-0127728.
 PR 30-MAR-2000: 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shinketsu RA, Leach M;
 XX WPI: 2000-602362/57.
 DR P-PSDB: AAB42665.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4044-4045; 5507pp: English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human OREF open reading frames 1 to 3161. The OREF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 CC osteoplastic; anticonvulsant; antihypertensive; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antitumor; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an OREF-associated disorder. The
 CC nucleic acids can be used to express OREF proteins in gene therapy. The
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, anti-inflammatory disease: to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other:
 SQ
 Alignment Scores:
 Pred. No.: 2.62e-39 Length: 2887
 Score: 1351.50 Matches: 269
 Percent Similarity: 94.74% Conservative: 1
 Best Local Similarity: 94.39% Mismatches: 8
 Query Match: 41.74% Indels: 8
 DB: 21 Gaps: 2
 US-10-054-935-2 (1-614) x AAC76874 (1-2887)
 QY 194 AlaSerGluGlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGlySer 213
 DB 5 GCCGGCAGGCGACATGGAAGATGATGAGAAAGCCCTCGGGGGTGGGGCGCTCG 64
 QY 214 GlyAlaSerSerGlnAlaAlaCysLeuLysGlnLeuLeuLeuLeuValLys 233
 DB 65 GGAGC-TCACGTCCGGCCGCTCCCTCAACAGATCCCTTCGCAATTGGACCTCATC 123
 QY 234 GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 253
 DB 124 GAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183
 QY 254 ArgAspThrLeuLeuAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 273
 DB 184 AGAGACACGCTCCCTTCGCTGATGAAAGATGAAAGGCGGATGCGTAAAGAG 243
 QY 274 AspAsnGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 293
 DB 244 GATACGAGAAAGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 303
 QY 294 ThrGluLeuSerGluLysIleLysLeuGluGlnProGluLeuSerGluThrSerGln 313

DB 304 ACAAGCTACTGTGAAATAAATTAACCTGAGTGCAGCGGAGCTTCGAGACATCCAG 363
 QY 314 ThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyLysArgLysSer 333
 DB 364 ACTCGCTCCCAAGCCCTTCATGTGGGGGAGGAAAGGACATAAAGGAAATCC 423
 QY 334 ProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerLys 353
 DB 424 CCATTGGAACTACAGAAAGAAAGACTCTGTTAAAGCTGGCTCTGAAATTTCAAAA 483
 QY 354 ValLysThrLysThrProLysHisSerProIleLysGluGluProCysGlySerLeuSer 373
 DB 484 GTCAAAACAAAACTCTTAAGCACTCTTAAGAGAACCTGGGTCTTATCT 543
 QY 374 GluThrValLysLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSer 393
 DB 544 GAAACGTGTTGTAACGTGAATTGAGAGCCCAAGAAACCCCGAGAAAGCCCGCTCTTCA 603
 QY 394 ValAspThrProProArgLysSerThrProGlnLysGlyProSerThrHisProLysGlu 413
 DB 604 GTGAGACACCCACCAAGACTCTCCACTCCCAAAAGGAGACCCAGACCCATCCCAAGAG 663
 QY 414 LysAlaPheSerSerGluLysLeuAspLeuProTyrLeuSerThrThrGluMetLysLeu 433
 DB 664 AAACCTTCACAGTGAATGAGATTTGGCTTCCACACAGAAATGATTTTG 723
 QY 434 CysArgTrpHisGlnProProProSerProLeuProLeuArgGluSerProLysLys 453
 DB 724 TGCTGTTGGACACCGCTCCCAACCGTACCATTTAGCGGAATCTCTCCAAAGAG 783
 QY 454 GluGluThrValAlaAlaArgCysLeuMetProSerSerValAlaGlyGluThrSerValLeu 473
 DB 784 GAGGAGACTGTAGCA-----AGTAAGCATAGAGAACTTGCTTT 825
 QY 474 AlaValProSerTrp 478
 DB 826 ---ATACCTTAGTGG 837
 RESULT 3
 ID ABR62818 standard; cDNA; 615 BP.
 XX
 AC ABR62818;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rat sequence differentially expressed in response to a hepatotoxin #725.
 XX
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KW differential expression; centrilobular necrosis; steatosis.
 XX
 OS Rattus norvegicus.
 XX
 PN W0200210453-A2.
 XX
 PD 07-FEB-2002.
 EF 30-JUL-2001; 2001WO-0523872.
 XX
 PR 31-JUL-2000; 2000US-222040P.
 PR 02-NOV-2000; 2000US-244880P.
 PR 11-MAY-2001; 2001US-290029P.
 PR 15-MAY-2001; 2001US-290645P.
 PR 22-MAY-2001; 2001US-292336P.
 PR 06-JUN-2001; 2001US-295798P.
 PR 13-JUN-2001; 2001US-297457P.
 PR 19-JUN-2001; 2001US-298884P.
 PR 09-JUL-2001; 2001US-303455P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

[illegible]

PS Claim 5; Page 4044-4045; 5507pp; English.

XX AACT4446 to AAC77606 encode the proteins given in AAB40237 to AAB4397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
CC osteoplastic; anticonvulsant; antiallergic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertensive; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy.
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;

Alignment Scores:

Pred. No.:	7,04e-16	Length:	2887
Score:	680.50	Matches:	134
Percent Similarity:	87.74%	Conservative:	2
Best Local Similarity:	86.45%	Mismatches:	12
Query Match:	21.02%	Indels:	7
DB:	21	Gaps:	2

US-10-054-935-2 (1-614) x AAC76874 (1-2887)

OY 324 ArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrPro 343
DB 2491 AGAGGGGTGGAGGGGTTAGACTTCTACCCCATTTGGAGACGAGAAAGAAAGACTCT 2432
OY 344 ValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysHisSerPro 363
DB 2431 GTTAAAGAGTGGCTCCCTGAAATTTTCAAAAGCAAAACAAACCTCTAGACACTCTCT 2372
OY 364 IleLysGluGluProGlySerLysSerLysThrLysLysArgGluLeuArgSer 383
DB 2371 ATTAAAGAGAACCTGTGGTCTTATCTGAACTGTTTGAACGTAATGAGAGC 2312
OY 384 GluGluThrProGluLysProArgSerSerValAspThrProProArgLeuSerThrPro 403
DB 2311 CAAGAAACCCAGAAAGACCCCGGTCTCAGTGGACACCCCAAGACCTCCACTGCC 2252
OY 404 GluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeu 423
DB 2251 CAAAGGAGACCCAGACCAATCCCAAGAGAAAGCCTTCTCAAGTGAATGAAAGATTGG 2192
OY 424 ProLysLeuSerThrThrGluMetLysLysArgLysArgLysGlnProProProSerPro 443
DB 2191 CGTACCTTTCACACAGAAATGATTTGTGCTTGACACCACTCCCAATCACCG 2132
OY 444 LeuProLeuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetPro 463
DB 2131 TTACCATTAAGGAAATCTCTGCCAAAGAGAGGAGACTGTAGCA----- 2087
OY 464 SerSerValAlaGlyGluThrSerValLeuAlaValProSerTTP 478
DB 2086 ---AGTAGCATAGAGAACCTTGTCTT---ATACCTAGTGG 2048

RESULT 5
AAK51958
ID AAK51958 standard; cDNA: 7713 BP.
XX
AC AAK51958;

XX 06-NOV-2001 (first entry)

DT Human polynucleotide SEQ ID NO 503.

DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

XX 09-AUG-2001.

PD 05-FEB-2001; 2001WO-US04098.

PF 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HSE-) HYSEQ INC.

PA Tang YF, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPT: 2001-476283/51.
DR P-PSDB: AAM78825.

Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -

PS Claim 1; Page 1809-1815; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, hematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX

Sequence 7713 BP; 2077 A; 1974 C; 2094 G; 1568 T; 0 other;

Alignment Scores:

Pred. No.:	2.3	Length:	7713
Score:	246.00	Matches:	182
Percent Similarity:	32.73%	Conservative:	88
Best Local Similarity:	22.06%	Mismatches:	289
Query Match:	7.60%	Indels:	266
DB:	22	Gaps:	33

US-10-054-935-2 (1-614) x AAK51958 (1-7713)

OY 11 AlaAlaAlaProAlaGlyLysAsnProGluGlnArgLeuAspTyrGluArgAlaAlaAla 30
DB 4345 GCTGCGGAGGACGACGAGCGG-----CCGCTCCGCTAGGGCGAGCTGTACGA 4392
OY 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuProArgHis 49

Oy		549	cgUvValThrSerPheProGluProAspA---ValGusSerLeuMetIleTh	567	
Dd		6514	TGCCACACCACCAAAAGAGGAGGAAGAGCAGACAAGAAGAGCCCAAGAACGA	6573	
Oy		567	rProPheLeuProValAlaAlaPheGlyAcArgProLeuProLysLeuThrProGlnAsn	587	
Dd		6574	GGCCTCAAAAGCCCAAGGTGGAGGAGAGAAAGAAACCTGCTGTCCAAAAGCCCAAAGANT-	6632	
Oy		587	heglJLeuProTrpLeuAspGluArGSerArGcYsaArgLeuGluIleGlnLysLysGInt	607	
Dd		6633	-----CCAAGATTGGAAGCCACAGA--AGGAAGAGCGTGAATAGAAAAAAGTCCC	668D	
Oy		607	hrProHisArg 610		
Dd		6682	CACCCACAGAGA 6692		
RESULT 6					
AAD30567					
ID	AAD30567	standard; cDNA;	5373 BP.		
XX	AAD30567;				
XX	21-MAY-2002	(first entry)			
DE		Human kinase polypeptide (PKIN-20) cDNA.			
KW		Human; kinase polypeptide; PKIN-20; gene therapy; Addison's disease;			
KW		leukemia; immune disorder; lymphoma; melanoma; developmental disorder;			
KW		acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;			
KW		asthma; Crohn's disease; rheumatoid arthritis; burstits; atherosclerosis;			
KW		cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;			
KW		cholestasis; cardiacnt; cardiovascular disorder; Niemann-Pick's disease;			
KW		lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;			
KW		drug screening; transgenic animal; antinflammatory; hepatotropic;			
KM		hypotensive; anti-HIV; enzyme; ss.			
OS	Homo sapiens.				
XX					
FH	key	Location/Qualifiers			
FT	CDS	1..4719			
FT		/tag= a			
FT		/product= "Human PKIN-20"			
FT	sig_peptide	1..111			
FT		/tag= b			
FT	mat_peptide	112..4716			
FT		/tag= c			
PN		/product= "Mature human PKIN-20"			
MO200208399-A2.					
31-JAN-2002.					
20-JUL-2001;	2001WO-US23092.				
21-JUL-2000;	2000US-22003BP.				
PR	28-JUL-2000;	2000US-222112P.			
PR	04-AUG-2000;	2000US-222831P.			
PR	11-AUG-2000;	2000US-224729P.			
(INCY-) INCYTE GENOMICS INC.					
(THOR/) THORNTON M.					
Thornton M,	Yue H,	Khan FA,	Gururajan R,	Hafalia AJA,	Wallia NK;
Patterson C,	Ramkumar J,	Gandhi AR,	Policky JL,	Baughn MR;	
Triibouley CM,	Bandman O,	Nguyen DB,	Lu Y,	Burford N,	Lai P;
Ding L,	Yao MG,	Elliott VS,	Reedipon SA,	Kearney LJ,	Lu DAM;
Greenwald SR,	Tang YT,	Xu Y,	Walsh RT,	Gietzen KU,	Yang J;
Hillman JL;					
WPI; 2002-206083/26.					
P-PSDB; AAE19162.					

PT New huan kinase polypeptide useful in diagnosis, prevention and
PT treatment of cancer, immune disorder, growth and developmental
PT disorder, cardiovascular disorder and lipid disorder -
XX
PS Claim 5; Page 194-196; 196pp; English.

CC The present invention relates to an isolated human kinase polypeptide
CC (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is
CC useful for diagnosing, treating and preventing cancer (e.g., leukemia,
CC lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency
CC syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's
CC disease, rheumatoid arthritis), a growth and developmental disorder (e.g
CC bursts, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a
CC cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial
CC infarction), and a lipid disorder (e.g., fatty liver, cholestasis,
CC Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of
CC drug screening techniques and to analyse the proteome of a tissue or cell
CC type. PKIN is useful for creating knockin humanised animals or transgenic
CC animals to model human diseases, in somatic or germ-line gene therapy, to
CC generate a transcript image of a tissue or cell type, for detecting
CC differences in the chromosomal location due to translocation, inversion,
CC etc., among normal, carrier or affected individuals, and as hybridisation
CC probes for mapping naturally occurring genomic sequences. PKIN is useful
CC in southern or northern analysis, dot blot or other membrane-based
CC technologies, in PCR technologies, in distick, plin, multifactor enzyme
CC linked immunosorbent (ELISA)-like assays and in microarrays utilising
CC fluids or tissues from patients to detect altered PKIN expression. The
CC present sequence is human PKIN-20 cDNA.

Sequence 5373 BP; 1037 A; 1636 C; 1812 G; 888 T; 0 other;

Alignment Scores:	
Pred. No.:	1.87
Score:	245.00
Percent Similarity:	33.07%
Best Local Similarity:	22.20%
Query Match:	7.57%
DB:	24
Gaps:	24
Length:	5373
Matches:	139
Conservative:	68
Mismatches:	199
Indels:	222

US-10-054-935-2 (1-614) x AAD30567 (1-5373)

Cy 11 A1A1A1AProLAAGLyG1yGAsnProGluGlnArgLeuAspTYRGLuaRGAlA1A1A1A 30
DB 1248 GCGGAAGCTCCACAGTGTCTGGA-----GCAGA 1274

QY 31 LeuclglyProgluAspctPro-----GlyalaIaIgluaIaHisPheLeuProArg 48
 1275 GAAAGTGGAGCTGAGCAGCAAGCACCAAGAGGCGCTGTGACGCCCCAC----- 1322
 Db

[illegible]

QY 69 ProSerProAlaIagIcysctylgIlylvsGIyArGclyleuleuLeuProAlaIaIaIa 88
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1368 GCGGGACAGGCTGCCAGAGATGTGTGAGGGGCACAGGCTCATATTGCCAGACGGATGGGCC 1427

Qy 89 ProGlyGlnGlnGlnSerTrp-----GlyGlySerValProLeuProCysPro 105
 1428 CCC-----AGCTGTGAGCCAGGTGACGAGTACAGTGACCTTACGGCAGGAGCT 1472

Oy		106	ProFroIaIThrLysGlnAlaGlyIleGly--GIYGLuProIaIaAlaIaGlyAlaGly	124
	:			
	:::			
Db		1473	TGACCGACATTCACCAGGGAGCTGCCTGCAGGCTGGCAGGGCTCGTAGGTACGAGGACAGA	15322

Oy. 125 CysSerProArgProLysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAla 144
 1533 GCGTCGACGAGGCCCA-----GGGGCGACGAGGCGA 1562
 Pb 1533 GCGTCGACGAGGCCCA-----GGGGCGACGAGGCGA 1562

QY 145 AAlaAlaLysGluProThrProTPrAlaIaGlyAspSylGlyClAlaAlaIaSerProAlaIa 164
||||| |||||
Db 1563 GCTCCTTCAGAGCGTACACAGAGGCCCCAGCAGAGAGAGCGGCCCCACGTAGCCAGACC 1622
||| :|||

```

OY 165 ThrAlaSerAspProAlaGlyProProPro-----174
Db 1623 GGCCTTGAGCTCCCTGAGAGAGAGCCGCGCTGCCAGAGGAGCTGAGGCGCAGT 1682
OY 175 ---LeuProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAla 193
Db 1683 GTCCTCCCTGAGCCG-----GACAGTGCAGCAGCTGCAGGAGACAGTGGGA 1727
OY 194 AlaSerGluGlyAlaArgTrpLysSerMetArgLysSer-----ProLeu 207
Db 1728 GCACAGCCTTG-AGGAG-TCGTCCAGGCCAAGACCATCCACACAGCCTCTGAGACCAAC 1785
OY 208 GlyGlyGlyGlyGlySerGlyAlaSerSerGlyAlaAlaCysLeuLysGluLeuLeu 227
Db 1786 GGAATGGGACCCCTGAGGGGTGGCTCTGAGAGGCCCAACTGAGAGAGAGAGTGGCGCC 1845
OY 228 LeuGluLeuAspLeuIleGluGluGluGluGluGluGluGluGluGluGluGlu 247
Db 1846 CTGGCAGAGCAGCTGGAGCAGGCGCCACAGCCACAGCCAGCCAGTGTAGAGAGAGGCTCTG 1905
OY 248 GluGluLeuLysSerGluAlaArgAspThrLeuLeuAlaArgIleGluArgMetGluArg 267
Db 1906 TGCACGCTGCAGAGAGAGAAACCGAGGCTGAGCGCGGAGCAGAGAGGCGCTAGAA-----1959
OY 268 MetGluLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGluGlu 287
Db 1960 GCAGAGCTGGCCAGAGCAGAGAGACAGACAGCGG-----CTGGAGGCT-----2004
OY 288 ThrGluGluArgGluGluThrGlu-----295
Db 2005 -----GAGCGCGGAGAGACGAGAGCAACTGGAGGCCAGCTGCCGACATCTCAGC 2058
OY 296 -----2058
Db 2059 TGGGTGAATGATGAGAGAGTCTCAAGAGCTAAGCTGAGCGCCCTGGCCACCAATGAGCA 2118
OY 302 LeuGluCysGluProGluLeuSerGluThrSerGluThrLeuProLysProPheSer 321
Db 2119 GAGGAGCTGGAGCTCTTGAGAGACGTAGGACACCGACAGCGCTCCCTGCCGCGCCACTGAA 2178
OY 322 CysGlyArgSerGlyLysGlyHisLysArg-----LysSerPro 334
Db 2179 ATGAGAGCCTCGGCGAGGCTGAGCTGACAGCTGAGAGCGCTGAGAGCGAGATCCGCGCAG 2238
OY 335 PheGlySerThrGluArgLysThrProValLysLys-----346
Db 2239 CAGGCGCTGCAGAGCGGCTGACACACAGTGCAGAGGCCAGCTGAGCGTGAAGCGCGCT 2298
OY 347 -----2298
Db 2299 CTGCAGAGGCCGAGAGACAGAGCCAGCGCCCTGCAACAGAGAGCTGCGCATGCTCGGGAG 2358
OY 357 LysThrProLysHisSerProIleLysGluGluProCysGlySerLeu-----372
Db 2359 GACCTCGGGGCGGAGGCGGAGTGCAGACACAGCCCTCAAACTCCCTGATCCCTCTCTG 2418
OY 373 -----SerGlu-ThrValCysLysArgGluLeuArgSerGluGluThrProGlu 388
Db 2419 TCCTCCGAGCTCAAGAGATTCCTGCAGAGGACCTTGCATCTCAGAGAGGCCACA 2478
OY 388 u-----388
Db 2479 AGGCATGAGAGAGCAGATCTGAGCGCGAGGCCAGCCAGCAGCTGCCATGGGGGCT 2538
OY 389 ---LysProArgSerSerValAspThrProArg-----399
Db 2539 GTGTTCCCAAGACACCCACTGCCACACAGCTCTACAGAGTCTTCTGCTAAGGA 2598
OY 400 -----LysSerThrProLysGlyPro-----407
Db 2599 TGGGCATGGGGCCTTGAGAGGCTTGAGTAATGCTGTCTCCCTCCAGCGCGCTCA 2658
OY 408 -----SerThrHisProLysGluLysAlaPheSerSerGluIleGlu 421

```

```

Db 2659 CACAGCTGCGCCCGGAGAGCTTCCATCCCGCA-----2692
OY 421 uAspLeuProTrpLeuSerThrThrGluMetTyrLeuCysArgTrp-----436
Db 2693 -----CCAAGTGTCTCCGCTGACCTCGGTGATGCTGGGCGCCGAGGCGCTG 2745
OY 437 -----HisGluProProProse 442
Db 2746 GGTGTGATGCTCGGCTTCTTGTACACAACTGTGCTCCACAGGCGCCAGCTG 2805
OY 442 rProLeuProLeuArgLysSer-----SerProLysGluGluThr 456
Db 2806 CCCGTCGCCCTGACCTCTCCGACAGAGCCCTGGAGATACCCCGAAGACAGCAGAGC 2865
OY 456 rVal-----AlaArgCysLeuMetProSerSerValAlaGlyGluThr 471
Db 2866 ACTGCTATAGAGGCTTCTGTGTCGTGCGCGGCGCTCAGAGTGTCCGGGCGGCGCAG 2925
OY 471 rValLeuAlaValPro 476
Db 2926 CGCGTGTGTCTGCC 2941

```

RESULT 7

AAZ22301 standard; cDNA; 9551 BP.

AAZ22301;

25-NOV-1999 (first entry)

cDNA encoding a human trichohyalin (TRHY) protein.

Human: trichohyalin; TRHY; protein; tissue structure; wound healing; terminal differentiation; epidermal tissue; proteinaceous gel;

breast implant; ss.

Homo sapiens.

US9598752-A.

28-SEP-1999.

14-FEB-1997; 97US-0800644.

30-APR-1993; 93US-0056200.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Kim I, Chung S, Park S, Steinert PM, Lee S;

WPI; 1999-561041/47.

P-PSDB; AAY30795.

Human trichohyalin useful for forming a proteinaceous gel that promotes wound healing -

Claim 1; Fig 3A-W; 126pp; English.

The present sequence encodes a human trichohyalin (TRHY) protein.

The protein is found in terminally differentiating epidermal tissue,

CC and is involved in forming the structural architecture of such

CC tissue. The trichohyalin protein is useful for forming a

CC proteinaceous gel which may then be used for healing wounds, or in

XX breast implants.

XX

Sequence 9551 BP; 2907 A; 1974 C; 2870 G; 1800 T; 0 other;

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Length: 9551

Matches: 162

Conservative: 82

Mismatches: 265

AA236989	AA236989 standard; cDNA: 7033 BP.
ID	AA236989 standard; cDNA: 7033 BP.
AC	AA236989;
XX	
XX	
DT	13-MAR-2000 (first entry)
XX	
XX	
DE	Human peripheral benzodiazepine receptor associated protein-1 cDNA.
XX	
KW	Human; peripheral benzodiazepine receptor associated protein-1;
KW	PRAK-1; peripheral benzodiazepine receptor; chromosome 17;
KW	central nervous system; immune system; gene therapy;
KW	PRAK-1 deficiency condition; endocrine system; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	198..5771
FT	CDS
FT	/tag= a
FT	/product= "peripheral benzodiazepine receptor
FT	associated protein-1"
PN	W09960117-A2.
XX	
XX	
PD	25-NOV-1999.
XX	
PF	06-MAY-1999; 99WO-FR01070.
PR	15-MAY-1998; 98FR-0006150.
PA	(SNFI) SANOFI-SYNTHELABO.
PI	Casellas P, Gallegue S, Jbilo O, Le Fur G;
DR	MP1: 2000-062455/05.
DR	P-PSDB: AAY53970.
PT	New PRAK-1 polypeptide that interact with peripheral benzodiazepine
PT	receptor, used to treat e.g. immune, central nervous or endocrine
PT	disorders -
XX	
PS	Claim 6; Page 28-38; 44pp; French.
XX	
CC	The present sequence encodes a human peripheral benzodiazepine receptor
CC	associated protein-1, designated PRAK-1. PRAK-1 interacts specifically
CC	with the peripheral benzodiazepine receptor. The PRAK-1 gene is
CC	localised on chromosome 17 in the q22-q23 region. The gene is
CC	associated with markers of pathologies of the central nervous system
CC	or immune system. The PRAK-1 nucleic acid is useful in gene therapy (of
CC	PRAK-1 deficiency conditions, e.g. disorders of the central nervous,
CC	immune or endocrine systems; as a source of diagnostic primers and
CC	probes (see AA236990-Z37023) and of antisense therapeutics; for
CC	recombinant production of the PRAK-1 protein; and for detecting allelic
CC	variants, mutations, deletions, insertions, loss of heterozygosity and
CC	gene rearrangements in the PRAK-1 gene. The PRAK-1 protein is used to
CC	raise specific antibodies and to screen for specific modulators
CC	(potential therapeutic agents). The antibodies are used as immunoassay
CC	reagents, e.g. for diagnosis of abnormal expression or accumulation of
CC	PRAK-1.
XX	
XX	
SO	Sequence 7033 BP: 1466 A; 2102 C; 2260 G; 1205 T; 0 other;
Alignment Scores:	
Pred. NO.:	4.78
Score:	236..00
Percent Similarity:	37.54%
Best Local Similarity:	25.62%
Query Match:	7.29%
DB:	21
US-10-054-935-2 (1-614) x AA236989 (1-7033)	
15 AlaAGlyGlyAsnProGluGlnArgLeuAspTyrGluArg---AlaAlaAlaLeuGlyGly 3	


```

Db 1496 CCT-GGACGACAGTCTGACGACATCGCGG-AGGTGGCCGACGCGGACAGCTGGAG 1553
Qy 341 sThProValylsLeuAlaProGluPheSerValylsThr----- 358
Db 1554 GTGGAGCATGACGAGCTCGGCTCAGCCTACGGGAGGAGGAGGTCCGGAGACTG 1613
Qy 359 -----ProLysHisSerPro-----IleLysGluGluProCysGly-----SerLe 372
Db 1614 CAGCAGGCCGAGCTGAAAGCCACAGAGAACATAGAGAGCGCGACGCTCTGGAGTCT 1673
Qy 372 uSerGluThrValCysLysArgGluLeu-ArgSerGluGluThrProGluLysProArg 392
Db 1674 ACCTTGATTCACAGCAGCCCGGCTTCAGAGCTCCAGAACAGTCCCGCAGCCAAAC 1733
Qy 392 eSerValAspThrProProArgLeuSerThrProGluLysGlyProSerThrHisProL 412
Db 1734 GAGCAGTTCAG--CTCTCTGGCAGACAGAACTCCAGCTTCCGCTGACCCGCGCC 1790
Qy 412 ySgIuLysAlaPheSerSerGluLeuLysAspLeuProTyLysSerThrThrGluMet 432
Db 1791 TTGGATCTGCT-----C 1802
Qy 432 yLeuLys-----ArgThrHisGln-ProProProSer----- 442
Db 1803 ACATCTGCTCGACTGGAGGAGCTTGAGAGACTGCCACACACCCCTGCTGCTGCTC 1862
Qy 443 ---ProLeuProLeuArgLysSerSerProLysLysGluGlu-----ThrVal 457
Db 1863 ATTCGACGCTTCGCGGGGCTGGCCCAAGACCTTGACCTCCGCGGCTCCCT 1922
Qy 458 AlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValProSer 477
Db 1923 GGGGCGCTGACCCCAAGTCTTCGAGCCTCCCTCCGACTCTCATCGGGGTCTCGA 1982
Qy 478 TrpArgAspHisSerValGluProLeuArgAspProAsnProSerAspLeuLeuAsn 497
Db 1983 AGGACGACCAAGAGGAGAGTCTCTCCAACTCTCCACCTCCGAGTCCATCCACAC 2042

RESULT 9
AAK52942
ID AAK52942 standard; cDNA: 4037 BP.
XX
AC AAK52942;
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2471.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN MO200157190-A2.
XX
PD 09-AUG-2001.
XX
PE 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSED INC.
PI Tang YT, Liu C, Drmanac RT, Yasundi V, Zhou P, Xu C, Cao Y, Ma Y;

```

```

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR WP: 2001-476283/51.
DR P-PSDB: AAM79809.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PS Claim 1: Page 4748-4749; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 4037 BP; 1173 A; 1058 C; 1233 G; 573 T; 0 other;

Alignment Scores:
Pred. No.: 3.18 Length: 4037
Score: 235.50 Matches: 183
Percent Similarity: 32.05% Conservative: 83
Best Local Similarity: 22.05% Mismatches: 291
Query Match: 7.27% Indels: 273
DB: Gaps: 22
US-10-054-935-2 (1-614) x AAK52942 (1-4037)

Qy 11 AlaAlaAlaProAlaGlyGlyAsnProGluInArgLeuAspTyArgAlaAlaAla 30
Db 638 GCTGCGGACGACGAGGCGG-----CCGCTCCGCTGTGGCGAGCTGTACGA 685
Qy 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuPArgHis 49
Db 686 GCGGAGGTCCGCGAGATGCCGCGCGGTCTGCTGCTGGG-----CGCGGC 733
Qy 50 ArgLysLeuLysGluPro-----GlyProProLeuAlaSerSerGlnGlySerPro 67
Db 734 GCGGCGTACGCTACGCTGAGAGACACCTGCTCGAGACATCGCGACGTGGCGCA 793
Qy 68 AlaPro-----SerProAlaGlyCysGlyGlyGlyGlyGlyGlyLeuLeuPro 84
Db 794 GCGGCTAGACGACGAGGCGCGGACGAGGAGGAGGCGGCGCGCGCGCTGGC 853
Qy 85 Ala-----GlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 97
Db 854 GCGCTTCCGCGACGAGGCGCGGCGCGCGCGGTGACCTCGACGAGAAAGGCGCGGCT 913
Qy 98 ---GlySerValProLeuProCysProProAlaThrLysGlnAlaGlyIleGlyGly 116
Db 914 GCAGGAGAGTGGCGGCTACCTGCGCGGCGCACCA----- 946
Qy 117 GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyArgAlaValLeuPro 136
Db 947 ---CCAGAGAGAGTGGCGGAGCTGCTCGGCAATCCAGGCGTCCGG----- 991
Qy 137 IleGlnThrGlySerIleValAlaAlaAlaLysGluProThrProTPAlaGlyAspLys 156
Db 992 ---CGCGCGGACGCGGACGAGTGCAGCGGACCGCGGCGGCGGCGGCGGCGGCGG 1048
Qy 157 GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro 176
Db 1049 GTCGCGCTGCGGAGATTCGCGCGAGCTTGAGAGCCACGCGGTGAGAGACGCTGCA 1108

```

```

QY 177 LeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
Db 1109 GTCCGAGGA-----GTGGTTCCGAGTGGAGCGCTGGACCGCTGGAGGAGCCAA 1159
QY 197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGly----- 211
Db 1160 GGT-----GAAACAGACAGCGCTATGCGCTCAGCGCAGAGAGGATTAAGTACCG 1210
QY 211 ----- 211
Db 1211 GCGTCAGCTGAGGCCAGACCAGACAGCTGAGGAGCTGAAAGCACCAGAGACTACT 1270
QY 212 -----GlySerGlyAlaSerSerGlnAlaAlaCysLeu-LysGlnI 225
Db 1271 GGAGAGCGAGCGCTCTGAGCTGGAGAGCGCTCATCAGGCGCAGACATGGCTCCACAGGA 1330
QY 225 eLeuLeuLeuGlnLeuAlaPleu---LleGlnGlnGlnGlnGlnGlnAlaLysGln 244
Db 1331 AGCCATTCAGCAGCGTGGAGCGCTGAGCTGAGACACCAAGTGGAGATGGCCGCCAGCT 1390
QY 244 uLysGluLleGluGluLeu----- 250
Db 1391 GCGAGAAATACCGAGACCTGCTCAATGTCAGATGGCTCTGGATATAGAGATAGCCGCTTA 1450
QY 250 ----- 250
Db 1451 CAGAAACTCTCGAAGGTGAAGAGTGTGGATTGGCTTGGCCCAATTCCTTCTGCT 1510
QY 250 ----- 250
Db 1511 TCCAGAAAGACTCCCAAAATTCCTGTGTCTCCACATCAATAAAGTGAAGGAAGGA 1570
QY 251 -----LysSerGluArgAspThrLeuAlaArgLleGluArgme 264
Db 1571 GAAGATCAAAAGTGTGAGAGTGTGAGAAAGAAACTGTGATT-----GTGGAGGAACA 1624
QY 264 tGluArgArgMetGlnLeuValLys-----LysAspAsnGlnI 277
Db 1625 GACAGAGAGAGACCCAAAGTGAAGTGAAGAGTACTGAAGAGAGATTAAGAGGCCAAGA 1684
QY 277 sGluArgHisLysLeuPheGlnGlyTrpGluThrGluLysGlu----- 292
Db 1685 GAGAGAGGCGCAAGAGAGAGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1744
QY 293 -----GluThrGluLeuSerGluLysI 300
Db 1745 AACAAAGTCTCCCCAGCAGAGAGAGCTGCATCCCAAGAGAGAGAGAGAGAGAGAG 1804
QY 300 eLysLeuGluCysGln-----ProGluLeuSerGluThrSerG 313
Db 1805 AAGAGAGAGAGCAAGTACCGGCTGAGGCCAAGTCCCAAGAGAGAGAGAGAGAGAG 1864
QY 313 nThrLeu----- 315
Db 1865 CCCAGCCGAGTCAAGTCCCTGAGAGAGCCCAAGTCTCCAGCAAGAGAGAGAGAGAG 1924
QY 316 -ProPro-LysProPheSerCysLysArgSerGlyLysGlyHis----- 329
Db 1925 ACCGCTGAGGAGCCCAAGTCCCAAGAGAGATGCAAAATTTCCAGCTGAGGTCA 1984
QY 330 -----LysArgLysSerProPheGlnSerThrGluArgLysThrProValLysL 346
Db 1985 AGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2041
QY 346 yLeuAlaProGluPheSerLys-----ValLysThrLysThrProL 360
Db 2042 CCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2101
QY 360 yHisSerProLleLysGluGluProCysLys-SerLeuSerGluThrValCysLysArg 379
Db 2102 CCAAGTCCCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161

```

```

QY 379 ----- 379
Db 2162 CCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2221
QY 380 -----GluLeuArgSerGlnLysThrProGluLysProArgSerVal--- 394
Db 2222 GCCAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2281
QY 395 -----AspThrProArgLysSerThrProGlnLysGlyProSerThrHisPro 411
Db 2282 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2341
QY 412 -----LysGluLysAlaPheSerSerGluLleGluAspLeuPro 424
Db 2342 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2401
QY 425 TyrLeuSerThrThrGluMetCysArgTrpHisGlnProProProSerProLeu 444
Db 2402 GTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2440
QY 445 ProLeuArgLys-----SerSerProLysGluGluThr 456
Db 2441 CCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2500
QY 457 ValAlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAlaValPro 476
Db 2501 AAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2560
QY 477 SerTrpArgAspHisSerValGluProLeuArgAspProAsn-----ProSerAsp 493
Db 2561 GAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2620
QY 494 LeuLeuGlnAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuGluLeuAsp 513
Db 2621 GCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2680
QY 514 GluLysArgArg-----LysArgTrpAspLleGlnArg---Lle 525
Db 2681 GAGAGATGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740
QY 526 ArgGluGlnArgGlyLeuGlnArgLeuGlnArgLeuArgMetLysLysLysGlyLleGln 545
Db 2741 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2800
QY 546 GluSerGluProGluValThrSerPheProGluProAspAsp-----ValGluSer 563
Db 2801 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2860
QY 564 LeuMetLysThrProPheLeuProValVal-AlaPheGlyArgProLeuProLysLeuTh 583
Db 2861 CCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2920
QY 583 rProGlnAsnPheGluLeuProTrpLeuAspGluArgSerArgCysArgLeuGluLleG 603
Db 2921 CCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2968
QY 603 nLysLysGlnThrProHisArg 610
Db 2969 AAAAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

RESULT 10

AAD33243
ID AAD33243 standard; cDNA; 3608 BP.

XX AAD33243;

DT 01-JUL-2002 (first entry)

DE Human secreted protein-encoding gene 7 cDNA clone HMVBP38, SEQ ID NO:17.

XX Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;

KW immune disease; autoimmune disease; anemia; multiple sclerosis; cancer;

KW rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;

KW	searcy syndrome; Gaucher's disease; neurological disease; cardiac arrest;
KW	Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
KW	cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
KW	thrombosis; ocular disorder; corneal infection; wound healing; cardiant;
KW	vascular; thrombolytic; cytostatic; nootropic; gene; ss.
XX	
OS	Homo sapiens.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	97..3018
FT	/tag- a
FT	/product= "Human secreted protein precursor"
FT	sig-peptide
FT	97..198
FT	/tag- b
FT	mat_peptide
FT	199..3015
FT	/tag- c
FT	/product= "Human mature secreted protein"
XX	
PN	WO200218435-A1.
XX	
PD	07-MAR-2002.
XX	
PF	17-JAN-2001; 2001WO-US01567.
XX	
PR	28-AUG-2000; 2000US-228084P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Komatsoullis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI	Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M,
PI	Ni J;
DR	WPI; 2002-281060/32.
DR	P-PSDB; AAE20799.
PT	
PT	Isolated nucleic acid molecule encoding a human secreted protein is
PT	used in preventing, treating or ameliorating e.g. Alzheimer's disease,
PT	cardio-/cerebrovascular disorders and multiple sclerosis -
XX	
PS	Claim 1; Page 420-421; 504pp; English.
CC	
CC	AAD33337-AA033280 represent cDNAs corresponding to 18 human secreted
CC	protein genes, and AAE20793-AAE20836 represent the proteins they encode.
CC	AAE20837-AAE20847 represent human secreted protein fragments. The genes
CC	and their corresponding secreted proteins are useful for preventing,
CC	treating or ameliorating medical conditions, e.g., by protein or gene
CC	therapy. Pathological conditions can be diagnosed by determining the
CC	amount of the new protein in a sample or by determining the presence of
CC	mutations in the new genes. Specific uses are described for each of the
CC	18 genes, based on the tissues in which they are most highly expressed,
CC	and include developing products for the diagnosis or treatment of
CC	immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
CC	infections, anemia, rheumatoid arthritis and multiple sclerosis),
CC	cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC	the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC	diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
CC	-Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
CC	cardiac arrest, tachycardia, angina and thrombosis), infections caused
CC	by bacteria, viruses and fungi and ocular disorders (e.g. corneal
CC	infections). Secreted proteins of the invention can also be used to
CC	promote wound healing, maintain organs before transplantation, support
CC	cell culture of primary tissues, modulate differentiation of embryonic
CC	stem cells, induce mesodermal tissue to differentiate in embryos,
CC	modulate mammalian characteristics (e.g. height and weight), modulate
CC	the catabolism, anabolism, energy storage, mental state, biorhythms,
CC	cardiac rhythms, reproductive potential, hormonal levels appetite,
CC	memory and stress. They can also be used as an additive to increase
CC	or decrease storage capabilities and nutritional content of food. The
CC	present sequence represents a human secreted protein-encoding CDNA of
CC	the invention.
XX	
Sequence 3608 BP; 738 A; 1077 C; 1199 G; 594 T; 0 other;	

[illegible]

QY 150 ThrProTrrAlaGlyAspLysGly-----GlyAlaAlaSerProAlaAlaThrAlaSer 167
 Db 356 -----AGGGCTCCAGAGGGGGCCAGCGCCCGGGAGCCCTCGAGAAAGGTGCC 406
 QY 168 AspProAlaGly-----ProProProLeuProLeuProGly 179
 Db 407 CACCGCTGGCGGGCGGAGATGACGTGGCGAAGAAGACCGCGCGCGCCGACACAGCCG 466
 QY 180 ProProProLeuAlaPro-ThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTr 199
 Db 467 CCGCGCGCGCGCGCGCGACCGCGCAGAGC---ACCGCGAGCAGCACCAGAACCCAG 523
 QY 199 PLYSerMetArgLysSerProLeuGlyGlyGlyGlySerGly----- 214
 Db 524 GCGGACGGCAAGGGCGGCGACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 582
 QY 215 -----AlaSerSerGluAlaAlaGlyLeu 223
 Db 583 CTCTCCCGCTCCCGCTTC 642
 QY 223 sglIle-----LeuLeuLeuGluLeuAspLeuIle----- 233
 Db 643 CAGCGCTGGCAGGGCGCTCAACTTCTCTTCTACCTCGCCCTGGTGGCGCGCGCGCTTT 702
 QY 233 ----- 233
 Db 703 CTGGGGGTGGTGGCTCCACACGTCCTGAGAGAGGTCCAGCGCGCGCGCGCGCGCGCA 762
 QY 234 -----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 247
 Db 763 GGAAGTCTCCCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 822
 QY 247 egluGluLeuLysSerGluArgAspThrLeuAlaArgIleGluArgMetGluArg 267
 Db 823 GCAAGTCTTGCACACCCACATTTTGAGACTTTTGAGTCCACTTGTGAAGAGCTCCCAACAT 882
 QY 267 gmetGluLeuValLysLys-----AspAsnGluLysGluArg---HisLys 281
 Db 883 ACAAGAGCTCCAGAGAAAGCTGTGAAGCAGAGGAGAGGTAGGTCAAGCGGATCAGCGA 942
 QY 281 sleuPheGluGlyThrGlu----- 289
 Db 943 AGTCTGCAGAACTCCAGATGAGATTCTCAAGACCTCTCGATGGATCCATGTGTGT 1002
 QY 290 -----GluArgGluGluGluGluLeuSerGluLysLysLysLysLysLysLys 305
 Db 1003 GAAGAGAGCCCGGAGGAGGAGCTTCCAGCTCCCTGAGAACGCGGAG---GAGCGGCT 1059
 QY 305 nProGluLeuSerGluThr-SerGluThrLeuProPolysProPheSerCysGlyArgS 325
 Db 1060 GAGGAGAGCTCCAAATCCATCAACGACACATCGCCA-----TCTCAGACAGAT 1110
 QY 325 erGlyLysGlyHisLysArgLysSer-----ProPhe----- 335
 Db 1111 CCAGAGAGGAGCCAGAGAGATCAATGACATGAGCAAGGCAAGGTTGCCCTCCGGAAGA 1170
 QY 336 -----GlySerThrGluArgLysThrProValLysLysLysLysLysLysLys 347
 Db 1171 ATCTGAGGGAGCAAGCAGATTGAAAGCCTTAAGAGACCTGTGAAGGATACAGAC 1230
 QY 348 -----AlaProGluPheSerLysValLys----- 355
 Db 1231 CTCAGCCAAAGTGGAGAGTGGAGATGAGAGCCCTGAGAGTACCTTCAGACATGGA 1290
 QY 356 -----ThrLysThrProLysHisSer----- 362
 Db 1291 GTCTGACATCTACACCGAGGTCCGAGCTGTAGCCCTCAAGCAGAGCAGAGGCTTT 1350
 QY 363 -----ProIleLysGlu---GluProCysGlySerLeuSerGluThrValCysL 378
 Db 1351 CAAGAGAGCGGCGCCACACAGAGAGCGGCTGCCCTGCAG-GCCCTCAGAGAGAGCTTCA 1409
 QY 378 ysArgGluLeuArgSerGluThrProGluLysProArgSerSerValAspThrProp 398

Db 1410 GGTCTGAGAGAGTCCGTCCTCCCGCTCCCGGAGAGAGATCCGAGACTGGAGAGAGCTCC 1469
 QY 398 roArgLeuSerThrProGluLysGlyProSer-----ThrHisProLysG 413
 Db 1470 GCCAGCTGAGTCCGATTCACACGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1529
 QY 413 Lu---LysAlaPhe-SerSerGluIleGluAsp---LeuProTyLysSerThrThrGlu 430
 Db 1530 CTTTGAAGCACTCCAGCAAAAGATCAGGAGACTGGACTCCAGGCTCCAGACGTGGAGG 1589
 QY 431 MetTyLys-----CysArgTrp-----HisGluPro-ProProSerProLeuPr 445
 Db 1590 ATGGAGGTCTCTCCATGAGGTGGCTTCTGGCGCGCAGACGAGAGGCTGGAGTCCCTCC 1649
 QY 445 OleuArgGluSerSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSe 465
 Db 1650 TGTCCAGAG---CCAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
 QY 465 rValAlaGlyGluThrSerValLeu---AlaValProSerTrp---ArgAspHisSerVa 483
 Db 1695 GCGCGCTGGAGAGGCTCGGTCCTCAGAGCAGACAGAGAGAGAGAGAGAGAGAGAGAG 1754
 QY 483 IgluProLeuArgAspPro 489
 Db 1755 GGAGCGTGGCGGAGACCA 1773
 RESULT 12
 AA160608/c
 ID AA160608 standard; cDNA: 3203 BP.
 AC AA160608;
 DT 22-OCT-2001 (first entry)
 DE Human polynucleotide seq ID NO 4597.
 KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 OS Homo sapiens.
 PN W020015312-A1.
 PD 26-JUL-2001.
 PE 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HXYE-) HXYEQ INC.
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM41452.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA038642-AA042213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Alignment Scores: 100% C, 99% G, 100% T; 0 other;
Pred No. 1

Record No.:	4-64	Length:	320
Score:	228.50	Matches:	151
Percent Similarity:	36.25%	Conservative:	77
Best Local Similarity:	24.0%	Mismatches:	199
Query Match:	7.06%	Indels:	204
DB:	22		

03-10-034-935-2 (1-614) x AA160608 (1-3203)

Db
3157 AGTCCCGCTCCGGGCGTCCT-----ATGTGTTGTAATAAATLGLNGLY 32

33 GlyProGluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgIvsen 52
|||||
QY ----- 313

3136 TCACCACATGGATCCAGCCCGCGCGCGCGCGCGCGCGCGCATGCGCTCGCGCC 3077

53 LysGluProCl₂ProProLeuAlaSerGlnGlySerProAlaProSerProAla 72
||||: |||
3076 :|||:|||||

3078 AAACAAAGGGGC-----TCCAAAGGGGGCCACGCGCGCGAGCCCCCCTCG 3032

[illegible]

90 gIyglngIgluGluSertmglucl...6... 2972

2971 GCGTGCCTGCCCCAGTCCCGGGGAGCC-----
:: ||| |||||:: ||| |||
-----LysValProLeuProCysProProAlaThr 109

110 LysGlnAlaGlyIleGlyGlyGluProAlaAlaGactgacgtctcgcggccgc 2921

[illegible]

130 LysTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaIleTyrGluPro 2861

2860 -----TCGGCCCAACAA--- 2870

150 ThrProTlaGlyAspIysGly-----GlyAlaAlaSerProAlaAlaThrAlaSer 167

2848 -----AGGGCTCCAAGGCGGCCACGGCGCGCGGAGCCCCCTCGAGAGAAGGGTCCC 2798

```

108 asproalagly-----ProProProleuProleuProgly 179
      |||:::|
2797 ccccccccc-----|||

```

180 ProProbofoult-5 -
CACCCTGCGGGCCGCAGTACGTGGCGAAGAAGCCGCCGCCGCCGACAGACCG 2738

[illegible]

199 p1ySse1nEtAr01vssErPro1o01.01.01
-----ACCCGACGACAGACCCCGAGAACCAg 2681

```

      :g-ysccrfoLencylglylgllyserGly----- 214
      :::: |||
      ::|||

```

[illegible]

Db 1767 -----16GCGCCAAACAAATATATACACCCGCACTGTCCAGAGCCCTC 1812
 QY 447 ArgGluSerProLysArgGluLunhrValAlaArgCysLeuMetProSerSerVal 466
 Db 1813 AGATGCTCATCAACAAGGAGAAAGAGACCACCTGGCTCCACTTATTCGGTCCGTAG 1872
 QY 467 AlaGlyGluThrSerValLeuAlaValProSer-----TTPArgAsp 480
 Db 1873 TGGCCCGTCTATCTGTGTAACCTTGACCTGTCCCTACATGCTCTTATATGGGTGAG 1932
 QY 481 -----HisSerValGluProLeuArgaspProAsnProSer 492
 Db 1933 GGTGGGGAGAGAGGAATATTGTCCAGTGCACACCC---AATGACCCCAATCCCAAC 1989
 QY 493 -----AspLeuLeuGluAsnLeu 498
 Db 1990 CTCAGGTGAGCGGAGCCTCTACTTGAAGGAGACTA 2022
 RESULT 14
 AAZ38858
 ID AAZ38858 standard; DNA; 3117 BP.
 XX
 AC AAZ38858;
 XX
 DT 17-FEB-2000 (first entry)
 DE Human Jurkat cell clone P2-2 AIM6 which affects TRRE activity.
 KW Human; Jurkat cell; tumour necrosis factor receptor releasing enzyme;
 KW TRRE; cytokine; TNF; identification; cytostatic; anti-inflammatory;
 KW cardiant; immunomodulator; antirheumatic; antibacterial; cancer;
 KW heart failure; cachexia; inflammation; endotoxic shock; arthritis;
 KW multiple sclerosis; sepsis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO958559-A2.
 PN 18-NOV-1999.
 PD 14-MAY-1999; 99WO-US10793.
 PF 14-MAY-1999; 99WO-US10793.
 XX
 PR 14-MAY-1998; 98US-0081385.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Gatanaga T, Granger GA;
 DR WPI: 2000-039067/03.
 XX
 PT Tumor necrosis factor receptor releasing enzyme modulators and
 PT polynucleotides
 PS Claim 2; Page 65-66; 106pp; English.
 XX
 CC The present invention describes isolated polynucleotides (A) comprising a
 CC sequence expressed at the mRNA level in Jurkat T cells and showing
 CC increased enzymatic activity for cleaving and releasing the tumour
 CC necrosis factor (TNF) receptor in genetically modified COS-1 cells
 CC expressing the receptor. The present sequence represents a specifically
 CC claimed clone which affects tumour necrosis factor receptor releasing
 CC enzyme (TRRE) activity. Methods from the present invention can be used to
 CC assess a disease condition associated with altered TRRE activity. The
 CC polypeptides, polynucleotides and antibodies can be used to decrease or
 CC increase signal transduction from a cytokine in a cell. The polypeptides,
 CC polynucleotides and antibodies may be used to treat heart failure,
 CC cachexia, inflammation, endotoxic shock, arthritis, multiple sclerosis
 CC and sepsis, and cancer.
 CC
 SQ Sequence 3117 BP; 723 A; 1001 C; 913 G; 480 T; 0 other;
 Alignment Scores:

Pred. No.:	6.78	Length:	3117
Score:	223.50	Matches:	185
Percent Similarity:	31.67%	Conservative:	76
Best Local Similarity:	22.45%	Mismatches:	210
Query Match:	6.90%	Indels:	353
DB:	21	Gaps:	42

US-10-054-935-2 (1-614) x AAZ38858 (1-3117)

Oy	12	AlaIaIProAlaIagIyGly	-----	AsnProIuGlnIarGleuAspTYrGluIarAla	---	28
Db	293	GCTGTCCAGCAGAGGTGGCTTCAGTAAAGTGCCCACTCTGTGTGGCTCCAGGGGCGGG				352
Oy	29	-----		-AlaAlaLeuGlyIPro		34

Db	353	CCCGAGAGGTGGAGAGGTGGGGGTGTACGTGACAGACAGCTGGGAGCAGACGCCAGGCCA				412
Oy	35	GluAspGluProGlyAlaIaIaGluAlaHisPheLeuProArGHisIarGlyLeuLys	---	53		
Db	413	GCTTCACACCCTTCAACATGAGAA	-----	-CTGCCACAGTCTGTCTCCCTTCACAGTGC		463
Oy	54	-----		-GluProIy		
				-----	ProIeuleuAla	60
Db	464	AACCAAGGGAGGCCGCATCCTGAGATGGAGTCCGACCTTACTATTAACCAACCTGAGGC				523
Oy	61	SerSerGlnGly	-----	GlySerProAlaProSerPro		71
Db	524	ACTGAGCGGGAGAAAGGGGGGGGCCCACTGACCTGGACCGCTATGTGGACCAATGATGCC				583
Oy	72	-----		-AlaGlyCysGlyGlyLysGlyIarGlyLeuLeuLeuProAlaGlyAlaAlaPro		89
Db	584	ACAGAAGGTGCAGCTGGAGTATAG	-----	-----	GGGGCCCCA	616
Oy	90	GlyGlnGlnGlu	-----	GluSerTyrGlyGlySerValProIeuProCysProProProAla		108
Db	617	GGCACCCCTGATTTCTTCCACGACGACCCAGAAACCCCAACAGTACTGCTCCCTGCA				676
Oy	109	ThrySGlnAlaGlyIle	-----	GlyGlyGluProAlaAlaIaIaGlyAlaGly	---	Cys 125
Db	677	ACCCTTCACGTGGCATTTGGGCCACACAGAGTAAACCGGAGGTCTTCGGGCGAGGCCACC				736
Oy	126	SerProIarPro	-----	-LysTyrGlnAlaValLeuProIleGlnThr		139
Db	737	GCCCCCAAAACCCGGTGGCTGCTTCCCTCCACAGAGAGCAGACGACGACGACGACCA				796
Oy	140	GlySerIeuValAlaIaIaIaIaIaLysGluProThrPro				151
Db	797	GCAGCAGCAGCAGCAGCAGCAGGAGCCCTTACCCAGATGCCGCTTTTGAAACTTCTA				856
Oy	152	-----		-----	TyrAla	153
		-----		-----		
Db	857	TTCCATGCAAGCAGCAACCCCTGGCAGCAACCCAGAGACTTTGGCTGCAAGCAGCTGGGCG				916
Oy	153	-----		-----		153
Db	917	ACTGGAGACGTCCACCTGGCTCACACAGCATGGACACCTTACCCTTCCCCCCCAACCC				976
Oy	154	-----		-GlyAspLysGlyAlaIaIaIaIaIaSerProAlaIaIaThr		165
Db	977	AGATATGACACCCAGAACTGCCAAGGGCCCTTCTGCAGAA	---	-CTCAGGCCCGGACGACGACG		1033
Oy	166	AlaSerAspProAlaGlyProProIeuProIeuProIeuProGlyProProIeu	-----	183		
Db	1034	GCTTACC	-----	-TCAGTTCAGATGCCCTTCCCGCGCGCTCCGCGCTCTCTAAGA		1087
Oy	183	-----		-----		183
Db	1088	GGGATATCCTGCCTCCACGCGCTGTGATGGGGCTGGCACCCAGCTGGGCGAGAGGCCAC				1147
Oy	183	-----		-----		183
Db	1148	TGGCAACCTGTCTTCACTACATCAGTGGCCCTGACAGACGCCGACCTGGCTTCCCTGGGGCA				1207

QY	184	AlaProthralaIatnrlaIaglyThrlleuAlaIaSerGluGly-----	197
Db	1208	GGCCATCTCTAAACCTCTGGAGATTCCGCTTGAGCTAGGAGATGCGACACTACTGCTGA	1267
QY	198	-----ArgTrpIysSerMetArgLysSerProLeuGlyGly 211	
Db	1268	TGGGAGAGACTAGCACCCCAATTGGCCGGAGACGAGAGAGCTCTCCATGGGACGAGGA	1327
QY	212	GlySerGlyAlaSer-SerGlnAlaAlaCysLeuAlnLeuLeuLeuGlnLeuAs 231	
Db	1328	GGCGATGAGGCGACGTAGACACAGGAGGACTGTGGGAG-----GNGCTACGGGCGG	1378
QY	231	pleuIleIgluGlnGlnGlnGlnGlnGlnAlaIaLysGluLysGluIleGluGluLeu 251	
Db	1379	AGTGTTCAGGACGACGCGACGAGCGCGCGGCA-----	1412
QY	251	SerGluArgAspThrIleuAlaArgIleGluArgMetGluArgMetGlnLeuVa 271	
Db	1413	-TCCAGAGAGCCCAATTGTCGAC-----CTGGCCAGAGAGCTGTGGAGCTGGC	1462
QY	271	ILysLysAspAsnGluLysGluArgHisLysLeuPheGlnIlyrGluThrGluArg 291	
Db	1463	CTCACTGCAGAAATGAAGAAGAT-----GGCAGTGGTTCTGAAGAAGAA	1504
QY	291	gGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGluTh 311	
Db	1505	GGCGAAAGATATATGGCTCACTACCAAG-----TGTGGGTGGAGTTTCTTGAGCC	1558
QY	311	rSerGlnThrLeu-ProProLys-----ProPheSer- 321	
Db	1559	TTCC-----TTAGCCACCAAGCGACGACGAGAAAGACAGTGGATGTACCTCATCAT	1612
QY	322	-----CysGly-----	323
Db	1613	CCCACTGTCTGTGCTGTGGCAACTGTGAGCCAACTGAGGACGCCAGGCTGAGGCT	1672
QY	334	-----ArgSerGlyLys-----GlyHisLysArgLys 333	
Db	1673	TGATGAGGACGGGAAGGCTTGTGAACAGAACCTGTGAGACACAAGCCATCAGTCATCGT	1732
QY	333	erProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerL 353	
Db	1733	CACCCGCGAGGGGTCCACCC-----GATATCCCGGACAGATGCTC-----A 1774	
QY	353	ysValLysThrLysThrProLysHisSerProIleLysGluGluProCysGlySerLeuS 373	
Db	1775	AGCTAGGCGGAGGACATGATGTCAACTTGGAGGGGAGGCTTCGTCGGGAAACCAAA	1834
QY	373	er-----GluThrValCysLysArgLLeuLeuArgSerGlnGluT 386	
Db	1835	CGAGCGGCCGACGCCGAGCCCTCATCTCATCCCAACCAAGCGGGGACATTTCATCGCCCC	1894
QY	386	hrProGluLysProArgSer-SerValAspThrProPheArgLeuSerThrProGlnLys 405	
Db	1895	TCCGCTTACTTCACATCATCCCACTTACCAAGACCACTGGCTCTCC-----CGT 1945	
QY	406	GlyProSer-ThrHisProLysGluLysAlaPheSerSerGluIleGlnAspLeuProTy 425	
Db	1946	GGCGCTACTAGCACACCCCTCTGAGCGGAGCTTT-----GAGTACATCTCC	1990
QY	425	rLeuSerThrThrGluMetTyrLeuCysArgTrpHisGlnProProProSerProLeuPr 445	
Db	1991	C-----TACACGCCGCCGCCCATCTCAGCCC	2017
QY	445	oleuArgGluSerSer-----	450
Db	2018	TGTGGGGAAGGCTGTGGCTCTTACTTAATGCCATCATATACACACGACCAATCCCTGC	2077
QY	451	-----ProLysLysGluGluThrValAlaArgCysLeuMetProSer-----	464
Db	2078	CCCTCTCTCCATCAGCGCTTAAGAGTGGCCCATCGACGCGTGTGCGGACTATACAGTGTGA	2137
QY	465	-----SerValAlaGlyGluThrSerValLeuAla-----	474


```

QY 167 erAspPro-----A 170
Db 1556 CCAGTCCATCATACCGCCCGAGGCTGAGCATGCGAGCGACTGGCGAGACTGGCAG 1615
QY 170 lagIProProProleuProleuProGlyProProProleuAlaProThrAlaThrAlaG 190
Db 1616 CAGCTCCAGCTCCAGACCCCGGGG-----CAGGATCCACTAGCTCTGATGCC----- 1662
QY 190 lYThrLeuAlaAlaSerLueGluArgTrpLysSerMetArgLysSerProleuGlyGlyG 210
Db 1663 -----AGAACTGCCGAATTACT----- 1680
QY 210 lYgIglySerGlyAlaSerSerGlnAlaAlaCysLeuLysGlnLeuLeuLeuGlnL 230
Db 1681 -----ACTTGATTTTCAG 1693
QY 230 euAspLeuIleGluGlnGlnGln-----GlnLeuGlnAlaLysGluLysGluIleG 248
Db 1694 TAGATGTCAAAAGAGATCATCAGAAAGTGTAATAGCTGTGAGATACAGACCCAGTG 1753
QY 248 lueGluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgGm 268
Db 1754 AAGAGATTGTAGATGTCAGATGGGAA-----GAAGACTACACGGCTTAA 1798
QY 268 etGlnLeuValLysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTrpGluT 288
Db 1799 TTCAAGTATGTGGAGATATAAACTCAAGGACCCAGCAGAGAG---AGCCAGCTCAGCCTC 1855
QY 288 hrGluGluArgGluGluThrGluLeuSerGluLysIleLysLeuGlu----- 303
Db 1856 TAGAAGAGATGGAACCTTCTGCAATCAAGATGAGGCTCAACTAGGCTCCAGGTCCTC 1915
QY 304 --CysGlnProGluLeuSerGluThrSerGlnThrLeuProProLysProPheSerG 323
Db 1916 CTATAGACCTGAACTTTCCTCCAGTGAAGACCAACACCACTCAGCT----- 1965
QY 323 lYArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArgLysThrP 343
Db 1966 --TGTGACTTCTTCTGGAGAGTGAATCTTCTCA-----GCCAGCAGCAACACCTC 2017
QY 343 roValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys----- 360
Db 2018 CTATCCCACTGAGCAGGCTGCTTCTTCTAGCCAGCTGATCTCCCTCCACCTC 2077
QY 361 -----HisSerProIleLysGluGluProCysGlySerLeuSerG 374
Db 2078 TGTGTTCTCCGAAAAAGACTGAATCTCCAGTCCAGCAAGAG----- 2118
QY 374 lUThrValLysLysArgLueLueArgSerGlnGluThrProGluLysProArgSerSerV 394
Db 2119 -----GCCACAGCTCAGACTCCAGATCCCTTAAGGAGGCGAG 2155
QY 394 alAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGlu- 413
Db 2156 AACCTTCTCCAGTCCAGCAAGAGTTCAGCTGAGCCACAGAGCC---CCTAAGGAGG 2212
QY 414 -----LysAlaPheSerSerGluIleGluAspLeuProTyrlLeuSerThrThrGluM 431
Db 2213 CTGAACCATCTGCACACCGACGAGGAGGCTCAAGTATCTCTGAAGTCCACTGAAGAGC 2272
QY 431 etYrLeuCysArgTrpHisGlnProProSerProleuProleuArgLysSerP 451
Db 2273 AAGAGGT-----CCAGCTCAGATTCAGAGCCCTCTATGAGGCGAGAAC 2317
QY 451 ro-----LysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaG 468
Db 2318 CTTCTCCAGCCAGCAGAGGCGCAGATTCAGAGCTCCAGAGCCCTTAAGAGAGTAGAAT 2377
QY 468 lYgIuThrSerValLeuAlaValProSerTrpArgAspHisSerValGluProLeuArg 488
Db 2378 CTTCAAGCCAGCAGATGGTC-----CCAGCTCAGC 2407

```

```

QY 488 sPProAsnProSerAspLeuLeuGluAsnLeuAspAspSerValPheSerLysArgHisA 508
Db 2408 TTCCAGAGCCA-----CTTAAGAGAGTTTCAGCTCAACCTCAG 2446
QY 508 lAlLysLeuGluLeu-----AspGluLysArgArgLysArgTrpAspIleG 523
Db 2447 CTCATTATGAGGTGACAGTTCACACACAGGTCAGAGATCAAGCTCAGCATTTGACATTGC 2506
QY 523 lNArgIleArgGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetYrLysLysG 543
Db 2507 CCAGTGTCAACATTCAACCTTTGACCTGGAGCTTACCATTTACTCCAGAACCCACTAAAG 2566
QY 543 lYIleGlnLysSer-----GluProGluValT 552
Db 2567 AGTTGACATTTCTACAGCCCTGAAGAGGCTATAGTTCTCCAAAGCACCCTTAAGTGTA 2626
QY 552 hrSerPhePheProGluProAspAspValGlu-SerLeuMetIleThr-----Pro 568
Db 2627 CA-----CTTCCACATCCAGACAGGTTCAAGATTACAGATTACACCTGACTCAAGCA 2680
QY 569 PheLeuProValAlaPheGlyArgProleuProLysLeuThrProGlnAsnPheGlu 588
Db 2681 CAGTTCAACCTGTGGACCTGGGG-----TTACCATCATCTCCAGAAATCCACTACGAGG 2734
QY 589 LeuProTrpLeuAspGluArgSerArgCysArgLeuGluIleGlnLysLysGlnThrPro 608
Db 2735 TTGAATTCTTCTCAACACGCGCAGGAGACCCCACTCAGCTCTTAAGAAAGTTGTAACCC 2794
QY 609 His 609
Db 2795 AAC 2797

```

Search completed: July 14, 2003, 22:16:50
 Job time : 412 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: July 14, 2003, 22:10:06 ; Search time 279 Seconds
(without alignments)
3474.922 Million cell updates/sec

Title: US-10-054-935-2
Perfect score: 3238
Sequence: 1 MTRSAVFKAAPAGNPE.....NSRCLTETQKQTPRTCK 614

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1105431 segs, 789497651 residues
Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+.p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10054935/runat_14072003_115026_24714/app.query.fasta_1.775
-DB=PubMed-Applications_NA -QW=Fastap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10054935 @CGN 1.1.57 @runat_14072003_115026_24714
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEJURY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1448.5	44.7	1529	10	US-09-925-302-212 Sequence 212, App
2	1055.5	32.6	615	10	US-09-917-800A-725 Sequence 725, App
3	625	19.3	365	9	US-09-918-995-30064 Sequence 30064, A
C	4	243.5	7.5	9025608	9 US-10-156-761-1 Sequence 1, Appli

5	234.5	7.2	3107	10	US-09-783-066-5	Sequence 5, Appl 11	
6	230.5	7.1	7065	9	US-09-991-496-115	Sequence 115, App	
7	230.5	7.1	7065	10	US-09-874-923-115	Sequence 115, App	
8	228.5	7.1	3169	9	US-10-037-270-713	Sequence 713, App	
9	226.5	7.0	15952	9	US-10-171-311-51	Sequence 51, Appl	
10	225	6.9	2108	10	US-09-962-832-225	Sequence 225, App	
11	223.5	6.9	3117	10	US-09-752-639-6	Sequence 6, Appl 11	
12	223.5	6.9	3117	10	US-09-984-198-6	Sequence 6, Appl 11	
13	222.5	6.9	8459	9	US-09-563-728A-31	Sequence 31, Appl	
14	221.5	6.8	6354	9	US-10-084-817-158	Sequence 158, App	
15	221.5	6.8	8459	10	US-09-817-913-8	Sequence 8, Appl 11	
16	221.5	6.8	8459	10	US-09-817-538-8	Sequence 8, Appl 11	
17	220	6.8	13555	9	US-09-764-891-7281	Sequence 7281, Ap	
18	219	6.8	9025608	9	US-10-156-761-1	Sequence 1, Appl 1	
19	218	6.7	6457	10	US-09-880-107-3389	Sequence 3389, Ap	
20	217	6.7	36778	9	US-09-860-846-5	Sequence 5, Appl 1	
21	217	6.7	36778	9	US-09-836-821-5	Sequence 5, Appl 1	
22	217	6.7	36778	10	US-09-861-289-5	Sequence 5, Appl 1	
23	217	6.7	37948	9	US-09-988-3848-5	Sequence 5, Appl 1	
24	217	6.7	38506	9	US-09-793-708-19	Sequence 19, Appl	
25	216	6.7	33023	10	US-09-880-107-3350	Sequence 3350, App	
26	213.5	6.6	2725	9	US-10-177-293-117	Sequence 117, App	
27	213.5	6.6	2725	10	US-09-962-436-40	Sequence 40, App	
28	213.5	6.6	4437	9	US-10-156-761-4189	Sequence 4189, App	
29	212.5	6.6	2732	9	US-10-175-523-193	Sequence 193, App	
C	30	211.5	6.5	3331	9	US-09-373-658-31	Sequence 31, Appl
C	31	210	6.5	57130	10	US-09-835-081-3	Sequence 3, Appl 1
32	208	6.4	14800	10	US-09-954-456-1601	Sequence 1601, App	
33	207.5	6.4	1855	10	US-09-864-761-19708	Sequence 19708, App	
34	207.5	6.4	1958	10	US-09-864-761-2927	Sequence 2927, App	
C	35	207.5	6.4	2001	9	US-10-156-761-3168	Sequence 3168, App
36	207.5	6.4	5001	9	US-10-156-761-2323	Sequence 2323, App	
37	206.5	6.4	2520	9	US-09-864-636A-573	Sequence 573, App	
38	206.5	6.4	2520	9	US-09-864-636B-577	Sequence 577, App	
C	39	206.5	6.4	3387	9	US-10-001-873-22	Sequence 22, Appl
40	206	6.4	2445	9	US-09-864-636A-571	Sequence 571, App	
41	206	6.4	2445	9	US-09-864-636B-575	Sequence 575, App	
42	206	6.4	2445	9	US-09-864-636A-579	Sequence 579, App	
43	206	6.4	2445	9	US-09-864-636B-581	Sequence 581, App	
44	206	6.4	2538	9	US-10-156-761-6189	Sequence 6189, App	
C	45	205.5	6.3	4826	10	US-09-772-304A-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-09-925-302-212
; Sequence 212, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925, 302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-212
Alignment Scores:
Pred. No.: 1.05e-94
Score: 1448.50
Percent Similarity: 95.64%
Best Local Similarity: 95.30%
Query Match: 44.73%
Length: 1529
Matches: 284
Conservative: 6
Mismatches: 7
Indels: 7


```
Db 363 CAGCGGNGCTGCCATTACAGACGGGCTCTCTCGGGGGGCCCAAGAGCTACGCC 422
Oy 152 TTPALAGIAspLySgLyGlyAlaAlaSerProAlaAlaThrAlaSerProAlaGly 171
Db 423 TGGGCTGGGAGACAGGGGGGGGCTCCCGCCAGCTCGACCGCTCGAGCCGGGGA 482
Oy 172 ProProLeuProLeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThr 191
Db 483 CCCCACACACTCTGCGCCGGGCCGCCACCCCTCGCGCCGACGACGCGGAGCC 542
Oy 192 LeuAlaAlaSerGlyAlaGlyTTPLySgSerLeuAlaGlySerProLeuGlyGlyGly 211
Db 543 CTGCGCGCCAGTGAGGCGCAGATGAGATGATAGAGAAAGCCCTCTCGGGGGTGGCGGC 602
Oy 212 GlySerGlyAla 215
Db 603 NGCTGGGAGCC 614

RESULT 3
US-09-918-995-30064
; Sequence 30064, Application US/09918995
; Publication NO. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30064
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30064

Alignment Scores:
Pred. No.: 1.01e-36 Length: 365
Score: 625.00 Matches: 118
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.16% Mismatches: 0
Query Match: 19.30% Indels: 0
Gaps: 0
DB: 9

US-10-054-935-2 (1-614) x US-09-918-995-30064 (1-365)
Oy 281 LysLeuPheGlnGlyTyrGluThrGluGluArgGluGluThrGluLeuSerGluLysIle 300
Db 8 GAGCTGTTTCAGGGCTTATGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67
Oy 301 LysLeuGluCysGlnProGluLeuSerGlnThrSerGlnThrLeuProProLysProPhe 320
Db 68 AAACCTGGAGTCCAGCCGAGACTTCCGAGACATCCGAGACTGCTCCGACGCCCTTC 127
Oy 321 SerGlyGlyArgSerGlyLysGlyHisLysArgLysSerProPheGlySerThrGluArg 340
Db 128 TCATGTGGGCGGAGTGAAGAGGACATAAAGAAATCCCATTTGGAGTGAAGAGAA 187
Oy 341 LysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLys 360
Db 188 AAGACTCTCTATAAAGAGCTGCTCTGATTTTCAAAAGTCAAAACAAACAACTCTTA 247
Oy 361 HisSerProLeuLysGlnGluProCysGlySerLeuSerGlnThrValCysLysArgLys 380
Db 248 CACTCTCTATTAAAGAGAGACCTGTGTCTTATCTGAACGTTTGTAAACGTGA 307
Oy 381 LeuArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArg 399
Db 308 TTGAGGAGCCAAAGAACCCAGAAAGCCCGGTCTTCAAGTGAACACCCACCAAGA 364
```

```
RESULT 4
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADATOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 5.3e-05 Length: 9025608
Score: 243.50 Matches: 132
Percent Similarity: 28.45% Conservative: 37
Best Local Similarity: 22.22% Mismatches: 214
Query Match: 7.52% Indels: 213
Gaps: 19
DB: 19

US-10-054-935-2 (1-614) x US-10-156-761-1 (1-9025608)
Oy 4 ArgSerAlaValPheLysAlaAlaAlaAlaProAlaGlyAsn-ProGluGlnArgLys 23
Db 2781266 AGATCTCCCGGCTACGACATCTCCGAAAGCCCGGGGAGACAGTCCGCTCGGCC 2781207
Oy 23 uAsPTrGluArgAlaAlaAlaLeuGlyGlyProGluAspGluProGlyAlaAlaGluAl 43
Db 2781206 GGGTGACGCGGAGACA-----GGACCGCGGATCGCGACATCTCTCTCG 2781159
Oy 43 aHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerG 63
Db 2781158 AGAACGGGAGAACCCGTCACCACTCTCGGCTCCACTCGCCGCCGCCAGGCGCG 2781099
Oy 63 nGlyGlySerProAlaProSerProAlaGlyCysGlyGlyLysGlyArgGlyLeuLeu 83
Db 2781098 TGGGCGCGCGCCAGTCTCTGACCGCGCGGAGTGGTACGCGCGGCG----- 2781053
Oy 83 uProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnSerTrpLysLysValProLeuP 103
Db 2781052 -CTGCGCGCGCGCGCGCGCGGGA-CGT-----CCGCGCGC 2781019
Oy 103 ocYsProProProAlaThrLysGlnAlaGlyLysGlyGlyGlnProAlaAlaAlaGlyAl 123
Db 2781018 GCACCTCTCTCGCGCGGAGCGCGCGCGCGCGGAGACCGCTCCCGCGCGCGCACCTC 2780959
Oy 123 aGlyCysSerProArgProLysTyrGlnAlaValLeuProLleGlnThrGlySerLeuVa 143
Db 2780958 CTCTGCTCTGACACCGCGCGCGCG-----GCCGGCGCGCGCGGTCTGCGCGGAGCCGCCG 2780905
Oy 143 lAlaAla-----AlaLysGluP 149
Db 2780904 CTGCGCGTGACCGCGGTGAGCATCCACCGGCTCAAGCGCGCTCTGAGAGACCGAC 2780845
```

[illegible]

```

Db      2779975  GACACGACGAGTGTGGGTGGACCTGCGCCGCGGAGGCGGAGACGAGCCGTCGCCGCGGCG 2779916
              ||| |||
OY      441 ----ProSerProLeuProLeuArgGluSerSerProLysLysGluGluThrValAlaAr 459
              ||| ||| ||| ||| ||| |||
Db      2779915  CAGCCCGGAGCCGACGCGCCCGCGGACATCCGCC----- 2779882
OY      459 gCysLeuMeProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrpAr 479
              ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2779881 -----CCGACCGGACCTACCTGCTGCACCGCGGTCTCGCGGCGCTCGGCTG-- 2779834
OY      479 gASPHisSerValGluProLeuArgAspProAsnProSer 492
Db      2779833 -----TCACCGCGCCGGAAGCTGCTGCCCTCG 2779807

RESULT 5
US-09-783-066-5
; Sequence 5, Application US/09783066
; Patent No. US20020142302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyun
; APPLICANT: Yang, Yonghong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Dirmac, Radoje T.
; TITLE OF INVENTION: No. US20020142302A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791C1P2D
; CURRENT APPLICATION NUMBER: US/09/783, 066
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: No. US20020142302A1 yet Assigned
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 5
; LENGTH: 3107
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)..(2850)
US-09-783-066-5

Alignment Scores:
Pred. No.: 5.89e-08 Length: 3107
Score: 234.50 Matches: 150
Percent Similarity: 34.06% Conservative: 85
Best local Similarity: 21.74% Mismatches: 272
Query Match: 7.24% Indels: 185
DB: 10 Gaps: 24

US-10-054-935-2 (1-614) x US-09-783-066-5 (1-3107).
OY      19 ProGluGlnArgLeuAspTryGluArgAlaAlaAlaLeuGlyGlyProGluAspGluPro 38
              |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      82 CCCGATGTCACGCTGGCTGCCCGCCGGGCGACCTAGAGCGTCCCACTCAGAACAAACAGC 141
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      39 GAlaAlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyPro-- 57
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      142 GGCAGTAGTCTCTGTCGTCACCTCCGCCACCCGCTCTGTCATGCTGGTGGCCCCAG 201
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      58 ---ProLeuAlaSerSerGlnGlyGly---SerProAlaProSerProAlaGlyCysGly 75
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      202 TCCCCACTCCGAGGATCCACAGGCTCCCTCGAAGTCTTCGCCCTCCATGATGCCA-TATGA 260
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Qy 76 GlyylslyArgGlyLeuLeuLeuProAlaGlyAlaAla----- 88
Db 261 GGGCCCTGGGCAAGGCGCTGAAGCCGAGCTCAGCCGCTCCCTCCAGCGCTGTCTATT 320
Qy 89 -----ProGlyGlnGlnGlnGlnSerThrGlyGlySerValProLeuPro 103
Db 321 CAGCACCCTTTGGCTGGACAGCAGCTGATGTCGTCATGGAGA-----CCC 371
Qy 104 CysProProProAlaThrIleGlyGlnAlaGlyIleGlyGlyGlnProAlaAlaAlaGlyAla 123
Db 372 TGTCTCTCTCCGCTCTGTGAGCTCGAGACGCTGGG-----CCCCCGCGCTCCGCGCGC 425
Qy 124 GlyCysSerProArgProLysTyrGlnAlaValLeuProIle-----GlnThrGlySer 141
Db 426 GGCCA-GGACCCCAACCCAGCCAGCCCGGAGCCTGTGAGCCTGCCACCATCGAGAGAG 484
Qy 142 LeuValAlaAlaAlaLysGlnProThrProThrAlaGlyAspLysGlyGlyAlaAlaSer 161
Db 485 CTCTGCAAGATCATCCACAGTGCAGAGCCCG-----GGCTCTCCAGATGGGCGCGCGCAG 541
Qy 162 ProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro----- 178
Db 542 GCACCTCTCTACTCTCCTCCCTAGAGGGCCCTCCAGCCATCCCTGGCTCCCGCTACC 601
Qy 179 -----GlyProProProLeuAlaProThr-AlaThrAlaGlyThrLeuAlaAlaLase 195
Db 602 TGCCCGAGGGGAGCTCCAAACCACTGTCGACAGGCCCAACCAAGCACACAGTATCATGC 661
Qy 195 rGlnGlyArgTyrIlePheSerMetArgLysSerProLeuGlyGlyGlyGly----- 212
Db 662 CACACCCCGA---GACCCTCTGAAACCATCTCTCTGTGTGGGGAGGAGCATCGAAGC 718
Qy 213 -----SerGly-AlaSerSerGlnAlaAlaLysLeu 223
Db 719 CGCCAGCCCAATCCGAGGGGTCGCCGAAGCGGTGGCTCTCCCGACACCCACCAACT 778
Qy 223 ysglnIleLeuLeuLeuLeuLeuAspLeuIleGlnGlnGlnGlnGlnGlnAlaL 243
Db 779 CCGAGGTG-----AAATGAACAGCTTTCGAGAACGCAAGAACAGCTGTGGAAG 829
Qy 243 ysglnLysGlnIle-----GlnGlnLeuL 251
Db 830 CAGAGGCTGAGCGCGAGCGGGGTCCCCACAGCTCCACTCCGCGCCCGCGAGGCGCTGA 889
Qy 251 ysserGlnArgAspThrIleLeuAlaArgIleGlnArgMetGlnArgMetGlnLeu 271
Db 890 GCTCGAGATAGTAGCTCAGCGCCGCGTGGAGAGAAACGACACCATCGAGGCTC 949
Qy 271 allysLysAspAsnGlu-----LysGlnArgHisLysLeuPheGlnGly----- 285
Db 950 AGAAGCGACGATGAGGCCATATTCGCCAAGCACAGCCGAGCGGTGGCAAAAGCGCT 1009
Qy 286 --TyrGlnThrGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 305
Db 1010 TCCTGACAGTGCAGCCCGGAGAGCTCTGG-----GAGCGGG 1048
Qy 305 lnProGlnLeuSerGlnThrSerGlnThrLeuProProLysProPheSerCysGlnArg 325
Db 1049 AAGCAGAGCGGAGAGGCGCATTCGCTCCAGTCCCTGGTGGAGAGCGCCGAGCGC 1108
Qy 325 erGlyLysGlyHisLysArgLysSerProPheGlySerThrGlnArgLysThrProVal 345
Db 1109 AGGGCCAGGCTGAGCCCACTCAGCGCC----- 1137
Qy 345 yslsLysLeuAlaProGlnPheSerLysValLysThrLysThrProLysHisSerProIle 365
Db 1138 -----AAGCAGTGCACCTTCTCCAGACACTGGCCCGCGTGC 1174
Qy 365 ysglnGlnProCysGlySerLeuSerGlnThrValCysLys----- 378
Db 1175 CCCACGAGGGGCTGGGGAATATACATCAGAGCGTCAAGCACTGATGCCCTTGAGCT 1234
Qy 379 -----ArgGlnLeuArgSerGlnGlnThrProG 388

```

```

Db 1235 CGCTGACGCGGACATGACAGAGCTCAGAGCAGCAGCAGCAGCGGCTCTGGCCCGCGC 1294
Qy 388 lLysProArgSerSerValAspThrProProArgLeuSerThrProGlnLysGlyPro 408
Db 1295 AGGGCCCGGATCCGCC-----CCACCACCTGTGCTGGGTGTCATCCCTGGCCCA 1345
Qy 408 erThrHisProLysGlnLysAlaPheSer----- 417
Db 1346 CGAGGGGCGCCAAAGCTGATCCCGCAGCCCGCGCGGCGAGTCCCGGACCCGCGCA 1405
Qy 417 ----- 417
Db 1406 GCCCGGGCGCGGGCCAGCCAGTCAACCCCGCAGCCGAAACACAGCGCGCCAGCGAGC 1465
Qy 418 -----SerGlnIleGlnLysLeuProT 425
Db 1466 TGCGGCTGGACCTTGACAGCGGTGCTTACGCCACCCAGAGGTAGACAGCTCCCGC 1525
Qy 425 yrlserThrThrGlnMetTyrLeuCysArgTyrPheGlnProProProSerProLeuP 445
Db 1526 ACCTG-----GCAAGTTCGCGCGAGAGGTCGCCGCTGC 1561
Qy 445 rLeuArgGlnSerSerProLysGlnGlnThrValAlaArgCysLeuMetProSer 465
Db 1562 AGAGCGCTCTTCATCTCTCGGGGAGAGAG-----CCCCCGAGG 1606
Qy 465 erValAlaGlyLysThrSerValLeuAlaPro-----SerThrArgAspHisSer 483
Db 1607 AGCCAGCGCGCGCGCGCGCTCATCGAGATCCCGTGGAGCGCTGCGAGATCCCGCG 1666
Qy 483 alGlnProLeuArgAspProAsnProSerAspLeuGlnLysLeuAspAspSerValP 503
Db 1667 CCGAGCGAGGAGAGCAGCGGAGCCCGCTGTGCTGAGAGATTCCTTGAGAGAGAGGCT 1726
Qy 503 heserLysArgHisAlaLysLeuGlnLeu----- 512
Db 1727 CTTCGAGAGGGGAGCGCGGGGTGGGTCTCTTACAGAGTGAAGCAAGCCTG 1786
Qy 513 --AspGlnLysArgArgLysArgTyrAspIleGlnArgIleArgGlnGlnArgIleLeu 532
Db 1787 AGGAGAGATGCGCCAAAGCGGGCGCACCTGAGCGGAGCGGAGCGGAGAGAGG 1846
Qy 532 lnArgGlnLeuArgMetTyrLysLysGlyIleGlnGlnSerGlnProGlnValT 552
Db 1847 AGGCGCGCGCGGAGAGAGTGGAGGAGGTGGAGAGAGAGAGCGGAGAGAGCGC 1906
Qy 552 hrserPhePheProGlnProAspAspValGlnSerLeuMetIleThrProPheLeuPro 572
Db 1907 CGAGGCTGGCCCAAGAG-----GAGGCCCGCGGGCCAGCGCCGCTGTGTCCG 1954
Qy 572 alValAlaPheGlyArgProLeuPro 580
Db 1955 CAGTCCGATGCGACTCCAGCCCT 1980

```

RESULT 6
 US-09-991-496-115
 ; Sequence 115, Application US/09991496
 ; Patent No. US20020169285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Webb, John R.
 ; APPLICANT: Dillon, David C.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Bhatia, Ajay
 ; APPLICANT: Coler, Rhea
 ; APPLICANT: Probst, Peter
 ; APPLICANT: Brannon, Mark
 ; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
 ; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
 ; FILE REFERENCE: 210121.420C9
 ; CURRENT APPLICATION NUMBER: US/09/991,496

: CURRENT FILING DATE: 2001-11-20
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 115
: LENGTH: 7065
: TYPE: DNA
: ORGANISM: Leishmania major and chagasi
: US-09-991-496-115

Alignment Scores:
Pred. No.: 2,666-07 Length: 7065
Score: 230.50 Matches: 156
Percent Similarity: 36.60% Conservative: 79
Best Local Similarity: 24.30% Mismatches: 242
Query Match: 7.12% Indels: 168
DB: 9 Gaps: 30

US-10-054-935-2 (1-614) x US-09-991-496-115 (1-7065)

QY 10 AlaAlaAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTyrGluArgAlaAla 29
DB 2616 GCTGACAGCGCGCTAGACACCGCCAGCAGCAGCGCGC-----CGAGCTGGA 2663
QY 30 AlaLeuGlyGlyPro-----GluAspGluProGlyAlaAlaGluAlaHisPheLeuPro 47
DB 2664 GGCACAGGTGGCGGCTGGCGCGCAGCAGCGGAGAGCTGACACAGCG-----CCT 2714
QY 48 ArgHisArgLysLeuLysGluProGlyProProLeuAlaSerSerGlnGlySerPro 67
DB 2715 AGACACCGCCAC-----GCGACA 2732
QY 68 AlaProSerProAlaGlyGlyGlyGlyLysGlyArgGlyLeuLeuProAlaGly 86
DB 2733 GCGCGCGGAGCTGAGAGCAGCGGGTGGCAGCGCTGGCGCGGA-----CCGCGACA 2783
QY 87 AlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 106
DB 2784 GCGCGCGGAGCTGAGCGCGGAGCGCGGAGCGGAGCTGAGCGCGCTGAGCAGCGCTGAGCAGCGCGCAC 2843
QY 107 ProAlaThrLysGlnAlaGly-----LleGly-----Gly 116
DB 2844 GCGACGAGCGCGCGGAGCTGAGAGCAGAGTGGCAGCGGCTGGCGCGGAGCGCGGAGCGCGCAC 2903
QY 117 GluProAlaAlaAla-----GlyAlaGlyCysSerProArgProLysTyrGln 132
DB 2904 GCGCGCGGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGCGACA 2963
QY 133 AlaValLeuProLleGlnThrGlnSerLeuValAlaAlaAla-----LysGlu 148
DB 2964 GCGAGCGCGC-----CGAGCTGAGGAGCAGAGTGGCAGCGGCTGGCGCGGAGCGCGGAGCGA 3017
QY 149 ProThrProThrAlaGly-----AspLysGlyGlyAlaAlaAlaSerProAla----- 163
DB 3018 GCGCGCGGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGCAC 3077
QY 164 AlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGlyProProProLeu 183
DB 3078 GCGAGCGCGCGCGAGCTGAGGAGCAGAGTGGCAGCGGCTGGCGCGGAGCGCGAGCGG 3137
QY 184 AlaProThrAla-----ThrAlaGlyThrLeuAlaAlaSerGlnGluArgTyrPlys 200
DB 3138 GCGCGGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGAGCGA 3197
QY 201 SerMetArgLysSerProLeuGlyGlyGlyGlyGlySerGlyAlaSerSerGlnAlaAl 220
DB 3198 GCGAGCGCGCGAGCT-----GAGAGCGAGGAGTGGCAGCGGCTGGCGCGGAGCGCGAGCGA 3251
QY 220 acylsLeuLysGlnIle-----LeuLeuLeuGlnLeuAspLeuIleG1 234
DB 3252 GCGCGCGGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGCAC 3311
QY 234 uGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 247

DB 3312 GCAAGACGCGCGGAGCTGAGGAGCAGAGTGGCAGCGGCTGGCGCGGAGCGCGGAGCGG 3371
QY 248 -----GluGluLeuLysSerGlnArgAspThrLeuLeuAl 259
DB 3372 GCGCGGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGCACGCA 3431
QY 259 argIleGluArgMetGluArgMetGlnLeuValLysLysAspAsnGlnLysGluArg 279
DB 3432 GCAAGCGCGGAGCTGAGGAGCAGCGGTGGCAGCGGCTGGCGCGGAGCGCGGAGCGGCGG 3491
QY 279 GHisLysLeu-----PheGlnGlyTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 298
DB 3492 CCAGAGCTGAGCGCGGAGCGCGGAGCTGAGCAGCGCGCTGAGCAGCGCGCGAGCA 3551
QY 298 uLysIleLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 318
DB 3552 GCGCGCGGAGCTGAGGAGCAGGT----- 3575
QY 318 sProPheSerCysGlyArgSerGlyLysGlnHisLysArgLysSerProPheGlySerTh 338
DB 3576 -----GCGAGCGCTGGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGG 3622
QY 338 rGluArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysTh 358
DB 3623 CGAAGCGCGAGGA-----GCTCAGCAGC 3646
QY 358 rProLysHisSerProLleLysGlnGluProGlyGlySerLeuSerGluThrVal----- 376
DB 3647 GCTAGACACCGCGGAGCGGAGCGCGGAGCTGAGGAGCAGCAGGTGGCAGCGGCTGGCGG 3706
QY 377 -CysLysArgGluLeuArgSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 394
DB 3707 CGAGAGCGGAGCAGAGG--CGCGCAGAGCTGGCGCGGAGCGCGGAGCGGAGCGGAGCGG 3765
QY 394 LasThrProProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGlu 414
DB 3766 CTAGACACCGCGGAGCGGAGCGCGGAGCTGAGGAGCAGAGTGGCAGCGGCTGGCGGCGG 3825
QY 414 sAlaPheSerSerGluLleGluAspLeuProTyrLeu----- 426
DB 3826 GACCGGAGCAGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 3885
QY 427 -----SerThrThrLysMetLysCysArgTyrPheGlnProProse 442
DB 3886 GACACGCGCGCAGCAGCAGCGCGGAGCTGAGGAGCAGAGTGGCAGCGGCTGGCGGAGC 3945
QY 442 rPro-----LeuProLeuArgLysSerProLysLysGlu 455
DB 3946 GCGGAGGAGCTGAGCAGCGCGCTGAGCAGCGCGGAGCGGAGCGGAGCGGAGCTG--GAGGC 4004
QY 455 uThrValAlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAla 475
DB 4005 ACGGTTGGGACGG-----CT 4019
QY 475 lProSerThrArgAspHisSerValGluProLeuAlaGAspProAspProSerAspLeu 495
DB 4020 GCGCGGAGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 4076
QY 495 uGluAsnLeuAspAspSerValPheSerLysArgHisAlaLysLeuLeu----- 512
DB 4077 GCAAGCGCTGAGCAGCGG-----ACGAGCAGCGCGCGGAGCTGAGGAGCAGCGGTCG 4130
QY 513 -----AspLysArgArgLysArgTyrAspIleGlnArgIleAr 526
DB 4131 ACGGTTGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 4178
QY 526 gGluGlnArgIleLeuGlnArgLeuGlnLeuArgMetTyrLysLysLysGlyIleGlnG1 546
DB 4179 CGCGGAGGAGCTGAGCAGCGGCTGAGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 4229
QY 546 uSerGluProGlnValThrSerPhePheProGluProAspAspValGlnSerLeuMetI1 566
DB 4230 GCTGAGGAGCAGAGTGGAGCGGCTGGCGCGGAGCGGAGCGGAGCTGAGCAGCGGCTGAG 4289

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3117 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 US-09-752-639-6

Alignment Scores:
 Pred. No.: 3,59e-07 Length: 3117
 Score: 223.50 Matches: 185
 Percent Similarity: 21.67% Conservative: 76
 Best Local Similarity: 22.45% Mismatches: 210
 Query Match: 6,90% Indels: 353
 Gaps: 42

US-10-054-935-2 (1-614) x US-09-752-639-6 (1-3117)

QY 12 AlaAlaProAlaGlyGly-----AsnProGluGlnArgLeuAspTyrGluArgAla---- 28
 DB 293 GCTGTCCAGCAGAGGTGGCTTCAGTAAAGTGGCCCAACTGTGTGATGGCTCCAGGGCGGG 352
 QY 29 -----AlaAlaLeuGlyGlyPro 34
 DB 353 CCCGAGCGTGGAGAGGTGGGGTGCATGACAGCAGCTGGCAGCAGCAGCAGCAGCAGCAGC 412
 QY 35 GluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgLysLeuLys--- 53
 DB 413 GCTTCACACCCCATTCACATGAA-----CTGCCACAGTCTGTCCCTCTACAGTGC 463
 QY 54 -----GluProGly-----ProProLeuAla 60
 DB 464 AACCAAGGGAGCCCGCATCTCGAGTGGAGTCCCGCACTTACTATAACCACTCGTAGGC 523
 QY 61 SerSerGlnGly-----GlySerProAlaProSerPro----- 71
 DB 524 ACTGAAGGGGAGAAAGGGGGGGCCACAGCTGGACCGGTATGTGGCAGCAATGATGCC 583
 QY 72 -----AlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro 89
 DB 584 ACAGAGGTGCGAGCTGAGGTAGG-----GCGGCCCA 616
 QY 90 GlyGlnGlnGlu-----GluSerTyrGlyGlySerValProLeuProGlySerProProAla 108
 DB 617 GGCACCCCTGAATTTCTTCCACGACGCCAAGAAACCCCAACCAACGATCCTGCCCCGCA 676
 QY 109 ThrLysGlnAlaGlyLys-----GlyGlyGluProAlaAlaAlaGlyAlaGly---Cys 125
 DB 677 ACCCTTCAGCTGGCATTTCCGCCACACAGGTAAACGGGAGGTCTTCGGCGAGGCCACAC 736
 QY 126 SerProArgPro-----LysTyrGlnAlaValLeuProIleGlnThr 139
 DB 737 GCCCCCAAAACCGGTGGCTGCTTCCCTCCACAGAACAGCAGCAGCAGCAGCAGCAACACA 796
 QY 140 GlySerLeuValAlaAlaAlaLysGluProThrPro----- 151
 DB 797 GCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 856
 QY 152 -----TyrAla 153
 DB 857 TTCCATGCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 916
 QY 153 ----- 153
 DB 917 ACTGGAGAGTCCACCTGGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 976
 QY 154 -----GlyAspLysGlyGlyAlaAlaSerProAlaAlaThr 165

DB 977 AGATATGAACCCAGAACTGCGCAAGGCCCTTCTCAGAGA---CTACAGCCCCGACGACG 1033
 QY 166 AlaSerAspProAlaGlyProProProLeuProGlyProProProLeu----- 183
 DB 1034 GCTACG-----TAGGTCAGATCCCTTCCCGCCGCTCCCGCCCTCTCTAAGGA 1087
 QY 183 ----- 183
 DB 1088 GGTATCTGCTCCACAGCGCCCTGATGGGGCTGGCACCAGCCTGGCAGAGGCCAC 1147
 QY 183 ----- 183
 DB 1148 TGGCAACCTGTCTCATATCATCTGCCCCCTGCACAGCCGACCTGCTCCCTGGGCA 1207
 QY 184 AlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGly----- 197
 DB 1208 GCCCATTCCTGACCTCTGGATTCCTCCGCTGGAGCTGAGGAGTGCAGCTACTGCTGA 1267
 QY 198 -----ArgTyrLysSerMetArgLysSerProLeuGlyGlyGly 211
 DB 1268 TGGGAGAGACTACACCCCAATGCGCGGAGCGAGAGGCTCTCCATGGGACGAGGA 1327
 QY 212 GlySerGlyAlaSer-SerGlnAlaAlaCysLeuLysGlnIleLeuLeuGlnLeuAs 231
 DB 1328 GGGCATGAGGGCATGAGCAGCAGGAGCTGTGGCAG-----GTGCTACGGGGCGG 1378
 QY 231 PheIleGluGlnGlnGlnGlnGlnGlnAlaLysGluLysGluIleGluLeuLys 251
 DB 1379 AGTATCTCAGACAGCAGGAGGAGCGCCCGGCA----- 1412
 QY 251 SerGluArgAspThrLeuLeuAlaArgIleGluArgMetGluArgMetGluLeuVa 271
 DB 1413 -TCCAGAGAGCCCAATTTGCTGACCC-----CTGCCACAGAAAGCTGTGAGCTGCG 1462
 QY 271 LysLysAspAsnGluLysGluArgHisLysLeuPheGlnGlyTyrGluThrGluGlu 291
 DB 1463 CTCACCTCAGATCAAAAGAT-----GGCAGTGGTCTTGAAGGAA 1504
 QY 291 gGluGluThrGluLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGlu 311
 DB 1505 GCGCAAAAGTGTATTTGGCTCCTCACTACCAAG-----TGGGGGTGAGTTTCTGAGCC 1558
 QY 311 rSerGlnThrLeu-ProProLys-----ProPheSer- 321
 DB 1559 TTCC-----TTACCCACCAAGCAGCAGCAGAGACAGACAGTGGATGCTACCTCAT 1612
 QY 322 -----CysGly----- 323
 DB 1613 CCACGTCTGTGCTGTGCGAAGCTGTGACCCCACTGAGGACAGCCAGCTGGAGCTCT 1672
 QY 324 -----ArgSerGlyLys-----GlyHisLysArgLys 333
 DB 1673 TGATGAGAGGGAGGAGGAGTGTGAACAGAACCTGCTGAGCAGACAGCATCAGTCATCG 1732
 QY 333 erProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSer 353
 DB 1733 CACCCGAGCGCGGTGCACCC-----GAATCCCGGAGCAGAGTCTC-----A 1774
 QY 353 yValLysThrLysThrProLysHisSerProIleLysGluGluProCysGlySerLys 373
 DB 1775 AGCTCAGGCGGAGGACATGATGCTCAAGTTGGAGGGGAGACCTTCCGTGGAAACCAA 1834
 QY 373 er-----GluThrValCysLysArgGluLeuArgSerGlnGlu 386
 DB 1835 GCAAGCGGCCAGCGCCGAGCCCATCATCATCCACCAAGCGGGGCACTTCAATCGGCC 1894
 QY 386 hrProGluLysProArgSer-SerValAspThrProProArgLeuSerThrProGluLys 405
 DB 1895 TCCCGTCTACTCCAAATCATCCCATCAGAGGACCACTCGCTCTCC-----CGT 1945
 QY 406 GlyProSer-ThrHisProLysGluLysAlaPheSerSerGluIleGluAspLeuPro 425
 DB 1946 GCGCTAGCTGACCAACCCCTGTGAGCGGAGCTT-----GAGCTACCTCC 1990

OY	425	rIeUsErThrGluMeTyriLeucysArGTpHISglInProPseRleUp	445
Db	1991	C-----TMACCGGGCCCCCATCTCAAGCC	2017
OY	445	oleAurGlUSeSer-----	450
Db	2018	TtTGCGGGGAAGGCTGTGGCCTTACTTCATTCGCATCATATCAACCAGAACCATCCCTCG	2077
OY	451	-----PrOlyslYsglUlthrValAlaArgCySleuMetProSeR--	464
Db	2078	CCCTCCCTCCATCCATCAGCGCTTAAGAGTGGCCATCCGACAGCTGCTCGGACTTAACAGTGTGA	2137
OY	465	-----SerValAlagLyglUthrSeRValIleuala-----	474
Db	2138	AGTAACCCCGCGTCTCTCTCTGTGTGATGGGGAGGCCACCCAGTAGACATCGACCCAG	2197
OY	475	-----ValProSeRTrParGAsPhISerValGI	484
Db	2198	GATCAACGTGGCGTCCGGTTCCAGCGACAATAATCCCTTGATGAGGGACCGTCCCTG--	2255
OY	484	uPTroLeuArgAspProAmnProSeRaspLeu-----GlusInleuAspas	500
Db	2256	-GCAGCTGAGATCCCACAAAGGCTGACTTGTTGGCACCCATTGGAGAGCACTTAAGAG	2314
OY	500	pSeRValPheSerLysrghISalalysleuglUleuAspglAllysArgrgylsArGT	520
Db	2315	CAGC-----CGGGAGAAGCAGAGGCAAGTGA	2341
OY	520	pASP-----	521
Db	2342	AGACCTGCTGACAGCGCCCTGCTCCAGCATTTTCCTGCTGGTGGCACCAACAGAGCT	2401
OY	522	----IlEglNarglEarglUglNarg-----IlEuglNarglEuglNleuArgme	538
Db	2402	GGCCCTGCACTGTCTGCAGCAATTCAGAGAGACATCTGGAAACGTGATTAAGCTGT	2461
OY	538	tTYrlYslYlsLySglyllEglnglUserglUpogluValIThrSeRpheheProgluPr	558
Db	2462	GCTTAACAACCCCTCGGGGCCCAACACATTCGCTGGCAACTTATCATACAGAGCTC	2521
OY	558	oASP 559	
Db	2522	TGAC 2525	
 RESULT 12 US-09-984-198-6 Sequence 6, Application US/09984198 Patent No. US20020106679A1 GENERAL INFORMATION: APPLICANT: Gatanaga, T. APPLICANT: Granter, G.A. TITLE OF INVENTION: Factors Altering Tumor Necrosis TITLE OF INVENTION: Factor Receptor Releasing Enzyme Activity, and Methods NUMBER OF SEQUENCES: 154 CORRESPONDENCE ADDRESS: ADDRESSEE: MORRISON & FOERSTER STREET: 755 PAGE MILL ROAD CITY: Palo Alto STATE: CA COUNTRY: USA ZIP: 94304-1018 COMPUTER READABLE FORM: MEDIUM TYPE: Diskette COMPUTER: IBM Compatible OPERATING SYSTEM: Windows SOFTWARE: FASTSEQ for Windows Version 2.0b CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/984,198 FILING DATE: CLASSIFICATION: PRIOR APPLICATION DATA:			

[illegible]

US-10-054 -935-2 (1-614) x US-09-984 -198-6 (1-3117)

OY 12 ALaAlaProAlaIeLyLyLy-----AsnProGluGlnArgLeuAspTryGluArgAla--- 28

Dd 293 GCCTGCCACAGCAGTGGCCCTCAGTAAGAATGGCCCCAACTCTGTGAATGGCTCCAGAGGC GGCG 352

OY 29 -----AlaAlaLeuGlyGlyPro 34

Dd 353 CCCGGAGCGTGAGGAGGTGGGGGTCTCAGTCAGCACAGCATGGCAGCGACAGCCAGCCAGCCCA 412

OY 35 GluAspLuproGlyAlaAlaGlnAlaHisPheLeuProArgHisArgLyLeuLyLeuLy--- 53

Dd 413 GCCCTCACACCATTTCACAATGGA-----CTGCCACAGTCTGTCCCTCTTACAGTGC 463

OY 54 -----GluProGly-----ProLeuAla 60

Dd 464 AACCAAGGGAGGCCCGCATCTCGAATGGAGTCCCGCATTAATAAACCACTGAGGC 523

OY 61 SerSerGlnLyLyLyLyLyLySerProAlaProSerPro----- 71

Dd 524 ACTGAAGCGGAGAGAAAGCGGGGGCCCAACAGCTGAGACCGATGCGACCAATGATGCC 583

OY 72 -----AlaIcylSeyLyLyLyLyLyArgLyLeuLeuProAlaGlyAlaAlaPro 89

Dd 584 ACAGAAGGTGCACACTGGAGGTAGG-----GCGGCCCA 616

OY 90 GlyGlnGlnGlu---GluSerTrpGlyGlySerValProLeuProOCysProProProAla 108

Dd 617 GGGACCCCTGAATTCTTTCCACAGCAGCAAGAAACCCCAAAACCAAGTCATCTGCCCTGCA 676

OY 109 ThrLySglnAlaIcylLyLyLyLyLyGlyGluProAlaAlaAlaGlyAlaGlyLy----Cys 125

Dd 677 ACCCTTCACACTGGCATTTGGCGCACCAAGSTGAACGGCAGGCTTCGGCGAGGGCCCAACC 736

OY 126 SerProArgPro-----LysTryGlnAlaValLeuProIleGlnThr 139

Dd 737 GCCCCCAAACCGGTGGTCCCTTCCTCCACAGAAAGCAGCAGCAGCAGCAACACCA 766

OY 140 GlySerLeuValAlaAlaAlaLySgluProThrPro----- 151

Db	797	GCACGACGACGACGACGACGACGACGACCCCTAACCCAGATGGCCGCTTTTGAAACTTCTA	856
OY	152	-----	TrpAla 153
Db	857	TTCCATGCGCAGCAACCACTTCGACGAAACCCAGACTTTGGCTGACGACGAGCTGGGCC	916
OY	153	-----	-----
Db	917	ACTGGGACAGTCCACCTGCGCTCACACACAGATGGACACCTTAACCTTCCCCCAACCC	976
OY	154	-----	-----
Db	977	AGATATGAACCCAGAACTGGCAGAGGCCCTTCTGGAGAA-----CTCAGCGCCGACGACGACC	1033
OY	166	AlaserAProAlaIglyProProProLeuProLeuProLeuProGlyProProProLeu-----	183
Db	1034	GCTACCC-----TCAGGTCCAGATCCCTTCCCGCGCGCTCCGCGCCCTCTCTTAAGA	1087
OY	183	-----	-----
Db	1088	GGGATATCGTGGCTCCACGAGCCCTGAGATGGGGCTGGACCAAGCTGGGACGAGAGCCAC	1147
OY	183	-----	-----
Db	1148	TGGCAACCTGTTCTACATCACTAGCCCTGACAGACCGCCACTGGCTCCCTGGGGCA	1207
OY	184	AlAProthralaIatrrAlaIglyThLeuAlaAlaserIgly-----	197
Db	1208	GCCCCATCCTGAACCTGTGGATTCCCGCTGAGACTGTGAGGAGTCCAGCTACATCGCTGA	1267
OY	198	-----	-----
Db	1268	TGGGAGAGACTAGCACCCAAATGGCCGGAGCAGAGAGGCTCCGCTATGGCCAGCAGAGA	1327
OY	212	glyserglyAlaser-SerGlnAlaAlaCysLeuLysGlnLeuLeuLeuGlnLeuAs	231
Db	1328	GGGCGATGAGGCGACATGAGCACAGGGAGACTGTGGGCAG-----GTGCTACGGGGCGG	1378
OY	231	pleuIIegIuGlnGlnGlnGlnGlnLeuGlnAlaIalysGlnLysGlnIleGlnIuLeuL	251
Db	1379	AGTATCCAGAGCAGCGGACCGAGCGCGCGGCA-----	1412
OY	251	sSerGlnaIaGspIhrLeuLeuAlaIatrgIleGlnIargMetGlnIaIargMetGlnLeuVa	271
Db	1413	-TCCAGAGGCCCAATTGTCTGACC-----CTGGCCCAAGAGGCTGTGGAGCTGGC	1462
OY	271	lLysIalysAspansGlnLysGlnIatrgIhIstLysIleuPheGlnIytrGlnItrGlnIuar	291
Db	1463	CTCACTCGAGATGCCAAGAT-----	-----GGCAGTGGTTCTGAAGAGAA 1504
OY	291	gGlnIuIthrGlnLeuSerGlnLysIleLysLeuGlnLysGlnProGlnLeuSerGlnIuTh	311
Db	1505	GCGGAAAGTATATATGGCTCACTACCAAG-----TGTTGGGTGGAGTTTCTGAGCC	1558
OY	311	rSerGlnIThrIeu-ProProLys-----	Proheser- 321
Db	1559	TTTCC-----TTAGCCACCAAGCGAGACGAGAAAGACAGTGGATGTACCCCTCATCAT	1612
OY	322	-----	-----
Db	1613	CCCAAGTGTCTGTGCTGTGGCAACTGTGGACCACTGAGGACCCAGCCAGCTGGAGGCT	1672
OY	324	-----	-----
Db	1673	TGATGAGGACGAGGAGGAGGTCTTGAACAGAACCTGTGTGAGCAACAAGCATCAGTCATCGT	1732
OY	333	eIrrProheGlySerThGlnIargLysIthrProValLysLysLeuAlaIProGlnIuPheSerL	353
Db	1733	CACCCGAGGCGGTCCACCC-----GAATCCCGGAGACAGATGCTC-----	-----A 1774
OY	353	ysValLysThIstIthrProLysIhIserProIleLysGlnIuIrrProCysGlySerLeus	373
Db	1775	AGCTCAGCGGAGGACATGAATGTCAAGATTGGAGGGGAGGACCTTCCGTGGCGGAACAA	1834

606

RESULT 13
US-09-563-78A-31
Sequence 31, Application US/09563728A
Publication No. US20030078216A1
GENERAL INFORMATION:
APPLICANT: Macleod, Alan R
APPLICANT: Li, Zoumei
APPLICANT: Besterman, Jeffrey M
TITLE OF INVENTION: Inhibition of Histone Deacetylase
FILE REFERENCE: 106101.229
CURRENT APPLICATION NUMBER: US/09/563,728A
CURRENT FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/132,287
PRIOR FILING DATE: 1999-05-03
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.1

901

Db 2135 GGGTGTCCCTCCATCCACAGCTGGCGCAGCCGCCACTGGGGGAGCCACTGCG 2194
 QY 516 rgarlysarqTTPasPILleqlnarqleargluegluarqllleuglnarqleuglnL 536
 Db 2195 CCCCCTGCCCCCAAGAGCCGCTCTGCAGCAGCTGTGCATCCAGCAGCAGCATCAGC 2254
 QY 536 euargmetYrlys---LysLysgLYlIleqlnuserLupProgluValTLThrsrPhep 555
 Db 2255 AGTTTCTGGAGAAACACAGCAGCAGTTCAGCAGCAGCAGCAGCAGCAGCAGTCA 2314
 QY 555 heProLupProasPaspValgluserLueMetleThrProPheLupProValAlap 575
 Db 2315 TCCCCAGCCAGCCGAG--- 2331
 QY 575 heglYarProLupProLysLupThrProgluasn--- 586
 Db 2332 ---CCAGCCCGCAGCCGAGAGCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2377
 QY 587 -----PheglLupProTLrPLeuasp-----GluA 595
 Db 2378 TCCGTGAGCAGCAGCTGTCTGTGAGCAGCAGCCCTGAGCCGCTGCCGGGCGAGAG 2437
 QY 595 rgserrargCysarqleugluIleqlnLysLysgluThrPro 608
 Db 2438 AGCCGACGACAGCCGCGCTGTGCAGGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGC 2478

RESULT 14

US-10-084-817-158
 : Sequence 158, Application US/10084817
 : Publication No. US20030119009A1
 : GENERAL INFORMATION:
 : APPLICANT: Susan Stuart
 : APPLICANT: Jed G. Nuchtern
 : APPLICANT: Sharon E. Plon
 : APPLICANT: Jason M. Shohet
 : TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
 : FILE REFERENCE: PA-0046 US
 : CURRENT APPLICATION NUMBER: US/10/084,817
 : PRIOR FILING DATE: 2002-02-25
 : PRIOR APPLICATION NUMBER: 60/270,784
 : PRIOR FILING DATE: 2001-02-23
 : NUMBER OF SEQ ID NOS: 365
 : SOFTWARE: PERL Program
 : SEQ ID NO 158
 : LENGTH: 6354
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc_feature
 : OTHER INFORMATION: Incyte ID No. US20030119009A1 1393778CB1
 US-10-084-817-158

Alignment Scores:
 Pred. No.: 1,04e-06 Length: 6354
 Score: 221.50 Matches: 174
 Percent Similarity: 32.23% Conservative: 90
 Best Local Similarity: 21.25% Mismatches: 254
 Query Match: 6.84% Gaps: 302
 Db: 9 Gaps: 35

US-10-054-935-2 (1-614) x US-10-084-817-158 (1-6354)

QY 9 LysAlaAlaAlaAlaProAlaGlyLys---AsnProgluLuarqleuaspTyrgLuarq 27
 Db 2503 CGGTGTCTGGCCACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
 QY 28 AlaAlaAlaLeu-----GlyLysProgluLuarqleuLysAlaAlaAlaAlaHisPhe 45
 Db 2563 GTTCAGAGGCTCTGTGAG 2622
 QY 46 LeuProArqHlsArqLysLeuLysgluProgluProPro----- 58
 Db 2623 GCTTACCCCATGAA-----GGTCTCTCCAGCGGAACCTGCGTGCCTACT 2667

QY 59 -----LeuAlaserSerglnly-----Gly 65
 Db 2668 GAAGCTGGGAAGACTGGAGAGTGGTGGCGCTTCTTACCAAGTCAAGCCGCTGTGCAGGT 2727
 QY 66 SerProAla-----ProSer 70
 Db 2728 GAGCCGAGAGAGAGAGAGATGATGCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2787
 QY 71 ProAlaGlySerglyLysLysgluArgly----- 80
 Db 2788 GCAGCTGGCTGGAG 2847
 QY 81 ---LeuLeuLupProAlaGlyLysAlaAlaProglu----- 90
 Db 2848 AGAGAAATTCACAGCTGTGAG 2907
 QY 91 -----GlnGlnGlu----- 93
 Db 2908 GGAAGCTCGGGCCGCTGTGAG 2967
 QY 94 -----GluSerTrpLysLysSerValProLupProCysPro----- 105
 Db 2968 AGAGGCAAGGCTGTGAG 3027
 QY 106 -----ProProAlaThrLysGlnAlaGlyLleGlyGlyGlu---ProAla 119
 Db 3028 GATGAG 3087
 QY 120 AlaAlaGlyAlaGlyCysSerProArqProLysLysLysLysLysLysLysLysLysLys 3139
 Db 3088 GAAGCTGCAGCTGGA----- 3102
 QY 140 GlySerLeuValAlaAlaAlaLysgluLupProThrProTLrPLeuLysLysLysLysLys 3159
 Db 3103 -----GAGGTGACACCCGAGCGGAA 3123
 QY 160 AlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLupProLupProglu 3179
 Db 3124 GCTGAAAAAGCTGGAG 3171
 QY 180 ProProProLupAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluLysLysLys 3199
 Db 3172 -----CAAGCTGCCAAGAGAAAGAACTGCTGGAGAGACAG 3207
 QY 200 LysSer----- 201
 Db 3208 AATAGCTGAGTTCAACCAACCTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3267
 QY 202 -----MetArgLysSerProLupGlyGly 209
 Db 3268 GCTCAGAGAACAGCATGAG 3327
 QY 210 GlyLysLysSerGlyAlaSerSergln-AlaAlaCysLeuLys----- 223
 Db 3328 GAAGCAGCAGCAG 3387
 QY 224 -----GlnLleLeuLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 3447
 Db 3388 CAGGAGACCAATGCGCCAGCTCCAGAGCCAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAG 3447
 QY 235 nglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3507
 Db 3448 GAAAG 3507
 QY 255 pThrLeuLeuAlaArgLleGluLysLysLysLysLysLysLysLysLysLysLysLysLys 3567
 Db 3508 CATGCGCCCTCAG 3567
 QY 275 nglnLysgluArqHlsLysLeuPheGlnGlyLysLysLysLysLysLysLysLysLysLys 3624
 Db 3568 GAGTCTGAGCGTGTCCAGAGATTAAGCT---GAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3624

Oy	295	uLeuSerGluLysIleuGluGlyuGlyGlnProGluLeuSerGluThrSer-----G	313
Db	3625	AGACGTAGAGCGCTGAAACAGAGATTGGAGACACGCTGGATTCCAGAGCTGCCACCA	3684
Oy	313	nThrLeuProProLysProPheSerCysGlyArgSerGlyLysGlyHisLysArgLysSe	333
Db	3685	GGAGCTC-----AGTCA-----AACTGGACCA	3708
Oy	333	rProPheGlySerThrGluArgLysThrProValLysLysLeuAlaProGluPheSerL	353
Db	3709	GGAGGTG---AACATCTCTGAAGAGACCCCTGAGAGAGAGGCCAACAGCCACGAGGCCCA	3766
Oy	353	sValLysThrLysThrProLysHisSerProIleLysLuuGluProCysGlySerLeuSe	373
Db	3766	GATCCAGAGATGAGCAGCAGACCTCCACAGCCCTGGAGAGAGCTGGCGGAGCAGCTGA	3825
Oy	373	rglThrVal---CysLysArgGluLeu---ArgSerGlnGluThrProGluLysProAr	391
Db	3826	GCAACACCAAGCGGGTGAAGACCAACTCTGAGAGAGCAAGCAAGACTCTGGAGAACGACG	3888
Oy	391	gSerSerValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHisPr	411
Db	3886	GGGGAGAGCTGCCAACAGAGTGAAGGTCTGCTGAGAGGCAAAAGGGAGCTGGAGACAC-	394
Oy	411	oLysGlnLysAlaPheSerSerGluIleGluAspLeuProLyr--LeuSerThrThrGlu	431
Db	3944	-AACCGCAGAAAGATGAGAGGGCAGCTGCAGAGAGCTGCAGGTCAAATTCCAAAGGAGGA	4002
Oy	431	etYrLeu-----CysArgTrpHis-----	437
Db	4003	GGCGGTGGCAGACAGCTGCCGCCACMAAGTCAACCAAGTGCAGGTGGAGCTGGACACACT	4066
Oy	438	-----GlnProProSerProLeuProLeuArgLysSerProL	452
Db	4063	GACCGGGCTTCTCAGCCAGTCCGACAGACAGTCCAGCAAGCTCACCAAGAGACTTCTCCGC	4122
Oy	452	ysLysGlu-----GluThrValAlaIaArgCysLeu-----	461
Db	4123	GCTGAGAGTCCAGCTGCAGACACTGCAGAGCTGCTGCAGAGAGAGAAACCGCACMAACT	4182
Oy	462	-----MetProSerSerValAlaGlyGluThrSerValLeuAlaValProser----	477
Db	4183	GAGCTTGAGCACCAAGCTCAAGCAGAGTGGAGAGCAGAGA-----AGAAATTCTTCCGGGA	4233
Oy	478	-----TrpArgAspHisSerValGluProL	486
Db	4237	GCAGCTGAGAGAGAGAGAGGCCAACACCAACTGTGAGAAAGCAGATCG-----CCAC	4299
Oy	486	euaArgAspProAsnProSerSerPheLeuGluAsnLeuAspSerVal-----	502
Db	4291	CCTCCATGCCCAGGTG-CCGCACATGAAAAAGAAATGAGAGACAGTGTGGGGTCTCCG	4349
Oy	503	-----PheSerLysArgH	507
Db	4350	AAACTGCTGAGAGAGGTGAAGAGAACTCCAGAAAGACCTGAGGGCCTGAGACGCGC	4405
Oy	507	IsaLalys-----	509
Db	4410	ACGAGAGAAAGTGGCCGCTACGACCAAGCTGGAGAAACCAAGACGGGGCTGCAGCAG	4466
Oy	510	-----LeuGluLeuAspGluLysArgArgLysArgTrpAspIleGln	524
Db	4470	AGCTGAGACACTGCTGTGGTGCACCTGCACCAACGCCGACAGGCCCTGCAACCTGGAG-	4528
Oy	524	rgIleArgGluGlnArgIleLeuGlnAlaArgLeuGlnLeuArgMetYrLysLysLysGly	544
Db	4529	-----AACACACAGAAAGATTGACCCAGCTC-----C	4556
Oy	544	IeGlnGluSerGluProGluValThrSerPhePheProGluLProAspAspValGluSerL	564
Db	4557	TGGCGAGAGAGAAAGACCATCTCTGCGCAAGTATGCGAGAGAGCGGACCGGGCTAGAGGG	4611
Oy	564	eumetIleThrProPheLeuProValValAlaIaPheGlyArgProLeuProLysLeuThrP	584

```

Db      4617 AGGCCCGAGAGAAGACCAAGCAGCTCTGTGCTGGCCCGGCCCCCTGGAGGAACCATNG 4676
        :          :          :          :          :          :          :          :          :          :
QY      584 roginAsnPhgeliuEuproFtrpleuAspgluArgSerArCyasArgleuGl 601
        |||         |||||         |||||         |||         ::|||
Db      4677 AGCAGAAAGCGGAGTGTGAGGGCGTCACAACAGCAATTCGCAGACGAGATGAG 4729
        :          :          :          :          :          :          :          :          :          :
RESULT 15
US-09-817-913-8
; Sequence 8, Application US/09817913
; Patent No. US20020061860A1
; GENERAL INFORMATION:
; APPLICANT: Li, Zumei
; APPLICANT: Bonfils, Claire
; APPLICANT: Besterman, Jeffrey
; TITLE OF INVENTION: Inhibition of Specific Histone Deacetylase Isoforms
; FILE REFERENCE: 106101.145
; CURRENT APPLICATION NUMBER: US/09/817,913
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 8459
; TYPE: DNA
; ORGANISM: Human
US-09-817-913-8

Alignment Scores:
Pred. No.:      1.4e-06      Length:      8459
Score:          221.50      Matches:    163
Percent Similarity: 33.20%      Conservative: 94
Best Local Similarity: 21.06%      Mismatches: 282
Query Match:     6.84%      Indels:    237
DB:              10      Gaps:       33

US-10-054-935-2 (1-614) x US-09-817-913-8 (1-8459)
QY      4 ArgSeraIaValPheLysAlaIaLaIaLaProAlaGIyASnProGIuGLnArgLeu 23
        |||||         |||||         |||||         |||||         |||||
Db      361 CGGTCCACACC GCCGCCGCCGCCGCCGCGCTGGAGAGCGGGGC-----CACGCG--- 408
        :          :          :          :          :          :          :          :          :          :
QY      24 AspyrGIuaRgaIaLaIaLaLeuGIyLProGIuAspGIuProGIyAlaIaGIuLa 43
        ::         |||||         |||||         ::         |||
Db      409 --TGCGCGCGCGCGCTGGAGACCGCGCGCTCCACAGAGCGCCGCG--GCCCTTCGTA 462
        :          :          :          :          :          :          :          :          :          :
QY      44 HisPheLeuProAaRhHisArgLysLeuLysGIuProGIy----- 56
        |||         |||||         |||||         |||
Db      463 CCTTTCCACCCGCGCGCGAGGCGGCTTCGCCGCCGCGGGGCGGGGCGGGGTGGGCA 522
        :          :          :          :          :          :          :          :          :          :
QY      57 --ProPLeuAlaSerGIuGIySerProAlaIaProSerProAlaGIyCsgly 75
        |||         |||         |||         |||||         |||||
Db      523 CCGCAGCGAGCGGCGCCGCTTCGCCGCTCGGGGCGCGGCCGCCCGA-GCAGGTTCARCT 581
        :          :          :          :          :          :          :          :          :          :
QY      76 GIyLysGIyaRgLyLeuLeuLeuProAlaGIyAlaLaIaProGIyGLnglUGlnUser 95
        ::         |||         |||         |||||         |||||         ::         ::
Db      582 GCAGAAAGCAGCGAGCGGCTGTGTCACACTGTGGGTACTGCGGTCAATGAGACCTTGCC 641
        :          :          :          :          :          :          :          :          :          :
QY      96 TrpGIyGIySeVaIProLeuProCysTrpProProAlaTrpLysGln-----Alagly 113
        :          :          :          :          :          :          :          :          :          :
Db      642 GGCGAGGCTCGGCGCTTGACGCTGTACCCACGACCCCTCACCGTCCGTAATTGTARGT 701
        :          :          :          :          :          :          :          :          :          :
QY      114 IlegIyGIyLuProAlaIaLaGIyAlaGIyCysSer----- 126
        ::|||         |||||         |||||         |||
Db      702 GTTGGCGGAGATTGGAGCTCGTTGGAGACTATCGTTTCCGTGGAAATTTTAGACCATTTTC 761
        :          :          :          :          :          :          :          :          :          :
QY      127 -----ProArgProLysTyGIaLa 133
        :          :          :          :          :          :          :          :          :          :
Db      762 GAATCACTTAAGAGATGACATTCGTACGATGAGCTCCCAAGCATTCCAGATGAGACT 821
        :          :          :          :          :          :          :          :          :          :
QY      134 ValIeuProIIegIntHcILysIerLeuValaIaLaIaLysGIuProThProTrIPala 153
        |||         |||||         |||||         |||

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 11:49:17 ; Search time 867 Seconds
(without alignments)
11356.098 Million cell updates/sec

Title: US-10-054-935-1
Perfect score: 4372
Sequence: 1 cagctcctgcgccccccgcac.....aaaaaaaaaaaaaaaa 4372

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002: *
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT: *
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1478	33.8	1509	AA224879	Human secreted pro
2	837	19.1	1529	AA18193	Lung cancer associ
3	785.4	18.0	2887	AA18193	Human OREF ORF2429
4	722	16.5	910	AA193984	Human neuroblastom
5	694.4	15.9	893	AA194300	Human neuroblastom
6	670	15.3	903	AA194299	Human neuroblastom
7	645.8	14.8	904	AA193963	Human neuroblastom
8	551.6	12.6	615	ABK62818	Rat sequence diffe
9	510.2	11.7	542	ABK39516	DNA encoding lung

10	403.2	9.2	451	16	AA222113	Human gene signatu
11	380.2	8.7	2887	21	AA224879	Human OREF ORF2429
12	302	6.9	317	21	AA225273	Human secreted pro
13	274.8	6.3	590	24	ABK62109	Rat sequence diffe
14	176.2	4.0	211	23	AAH36707	Human colon cancer
15	134.8	3.1	211	23	ABV56446	Human prostate exp
16	107.2	2.5	565	24	ABK62830	Rat sequence diffe
17	79.2	1.8	114955	20	AAK53491	Human adenocine AI
18	78	1.8	114955	20	AAK53491	Human adenocine AI
19	72	1.6	1127	21	AAK02477	Human colon cancer
20	70	1.6	801	20	AAV73801	KSHV LDR terminal
21	68.6	1.6	612	22	AAH71471	Human cervical can
22	67.2	1.5	936	18	AAH95102	Exon 3 encoding am
23	64.8	1.5	2561	22	AAH26500	Rabbit low density
24	64.6	1.5	1614	22	AAH26499	Human low density
25	64.6	1.5	12425	22	AAH26495	Human low density
26	63.2	1.4	1614	22	AAH26499	Human low density
27	63.2	1.4	12425	22	AAH26495	Human low density
28	62.8	1.4	556	23	ABV40063	Human prostate exp
29	62.8	1.4	556	23	ABV40163	Human prostate exp
30	62.8	1.4	556	23	ABV42105	Human prostate exp
31	62.8	1.4	556	23	ABV43601	Human prostate exp
32	62	1.4	801	21	AAA30291	Rhadinovirus cis-
33	62	1.4	801	24	ABA33488	Kaposi's sarcoma-a
34	61.4	1.4	372	23	ABV37528	Human prostate exp
35	61.4	1.4	2307	24	ABK32842	DNA encoding human
36	61.2	1.4	266	23	ABV07596	Human prostate exp
37	61.2	1.4	4411529	22	AA199682	Mycobacterium tube
38	60	1.4	60	24	ABN49572	Human spliced tran
39	59.6	1.4	309	23	ABV44994	Human prostate exp
40	59.6	1.4	626	23	ABV06941	Human prostate exp
41	59.6	1.4	109519	22	AA508693	Microsomopora DNA
42	59.2	1.4	309	23	ABV44994	Human prostate exp
43	58.8	1.3	2388	22	AAE52196	DNA encoding membr
44	58.8	1.3	2388	24	ABK12568	DNA encoding cell
45	58.6	1.3	989	21	AAA02539	Human colon cancer

ALIGNMENTS

RESULT 1	
AA224879	AA224879 standard; DNA; 1509 BP.
XX	AA224879;
XX	02-DEC-1999 (first entry)
DE	Human secreted protein gene 69 clone HCEK38.
XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
OS	Homo sapiens.
XX	
PN	WO9947540-A1.
XX	
PD	23-SEP-1999.
XX	
PF	18-MAR-1999; 99WO-US05804.
XX	
PR	19-MAR-1998; 98US-0078563.
PR	19-MAR-1998; 98US-0078566.
PR	19-MAR-1998; 98US-0078573.
PR	19-MAR-1998; 98US-0078574.
PR	19-MAR-1998; 98US-0078576.

XX	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KV	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
XX	
PR	02-APR-1999; 99US-0127636.
XX	
PR	05-APR-1999; 99US-0127728.
XX	
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
P1	Shimkets RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
XX	
DR	P-PSDB; AAB42665.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
XX	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Claim 5; Page 4044-4045; 5507pp; English.
XX	
CC	AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerrary;
XX	antiprolitic; antiparkinsonian; nootropic; neuroprotective;
CC	osteopethic; anticoagulant; antihypertic; immunosuppressant;
XX	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
XX	antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC	antihypoid; and antianhaemic. The sequences can be used for determining
XX	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
XX	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
XX	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
XX	hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
XX	Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;
XX	
XX	Query Match 18.0%; Score 785.4; DB 21; Length 2887;
XX	Best Local Similarity 99.7%; Pred. No. 4.2e-150;
XX	Matches 797; Conservative 0; Mismatches 1; Indels 1; Gaps 1
QY	655 CGGCGAGGAGGAGGAGATGAAAGATAGAGAGAGAGCCCTCTGGGGGTGGTGCGGCT 714
DB	3 CGGCGGCGGAGGCGCATGATGAAAGACTATGAGAGAGAGCCCTCTCGGGGTGGTGCGGCT 62
QY	715 CGGAGGCTCCAGTCAGGCGCCCTGCTCAAAACAGATCCTTCTGTCGAATTGGACCTCA 774
DB	63 CGGAGG-CCTCAGTCAGGCGCCCTGCTCAAAACAGATCCTTCTGTCGAATTGGACCTCA 121
QY	775 TCGAACAAGCAGCAGCAGCAGCTGCAAGGCCAAGAAAGAGAGATGAGAGAGCTGAAGTCAG 834
DB	122 TCGAACAAGCAGCAGCAGCAGCTGCAAGGCCAAGAAAGAGAGATGAGAGAGCTGAAGTCAG 181
QY	835 AGAAGACAGCCTCTTCTGCTGATTTGAACGTATGAGAAAGCGAGTCAGCTGGTAAAGA 894

Dd		182	AGAGGACACAGCCTCTTCTGTGGATTTGAACGATGTGAAAGGCGATGCAGCTGTAAAGA	241
Oy		895	AGGATAACGAGAAAAGAACGCCACAAGCTGTTCACGGCTATGAACCTGAGAGAGAGAG	954
Dd		242	AGGATTAAAGAGAAAAAGAAAGGCAACAAGCTGTTCACAGGCTATGAAACTGAAGAGAGAGG	301
Oy		955	AAACAGACCTATTCGAGAAAATTTAACGTGGAGTGCCACCGGAGCTTCCGAGCATCCC	1014
Dd		302	AAACGAGACTATCTGAGAAAATTTAACGTGAGGTGCGACCGGAGCTTCCGAGCATCCC	361
Oy		1015	AGACTCTGCCCCCACAGCCCTTCTCATTGTGGGCGAGGTGGAAGGAGCATAAAGGAAT	1074
Dd		362	AGACTCTGCCCCCACAGCCCTTCTCATTGTGGGCGAGGTGGAAGGAGCATAAAGGAAT	421
Oy		1075	CCCCATTTTGAAGATACAGAAAAGAAAGACTCCTGTTAAAAAGCTGCTCCTGAATTTCAA	1134
Dd		422	CCCCATTTTGAAGATACAGAAAAGAAAGACTCCTGTTAAAAAGCTGCTCCTGAATTTCAA	481
Oy		1135	AAGTCAAAACAAAAAACCCTTAAGCACTCCCTATTAAAGAGAACCTGTGGTTCCTAT	1194
Dd		482	AAGTCAAAACAAAAAACCCTTAAGCACTCCCTATTAAAGAGAACCTGTGGTTCCTAT	541
Oy		1195	CTGAACACTGTTTGTAAACGTGAATTTGAGAGAGCCAGAAAACCCACAGAAAAGCCCGTCT	1254
Dd		542	CTGAACACTGTTTGTAAACGTGAATTTGAGAGAGCCAGAAAACCCACAGAAAAGCCCGTCT	601
Oy		1255	CAGTGGACACCCCAACCAAGACTCTTCACCTCCCCAAAAGGAGACCCAGCCATCCCAAG	1314
Dd		602	CAGTGGACACCCCAACCAAGACTCTTCACCTCCCCAAAAGGAGACCCAGCCATCCCAAG	661
Oy		1315	AGAAAGCCTTCTCAAGTAGAGATGAGATTTGCGGTACCTTCCACCAAGAAATGTAT	1374
Dd		662	AGAAAGCCTTCTCAAGTAGAGATGAGATTTGCGGTACCTTCCACCAAGAAATGTAT	721
Oy		1375	TGTGTCGTTGGCACACAGCCCTCCCCCATACGCTTACCAATTACGAGGAATCTCTCCAAGA	1434
Dd		722	TGTGTCGTTGGCACACAGCCCTCCCCCATACGCTTACCAATTACGAGGAATCTCTCCAAGA	781
Oy		1435	AGGAGGAGACTGTAGCAAG 1453	
Dd		782	AGGAGGAGACTGTAGCAAG 800	
RESULT 4				
AI193984/c				
ID	AI193984	standard:	cDNA; 910 bp.	
XX	AAI193984;			
XX	AAI193984;			
DT	13-NOV-2001	(first entry)		
DX	Human neuroblastoma expressed polynucleotide SEQ ID NO 59.			
KW	Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.			
OS	Homo sapiens.			
XX	WO20016719-A1.			
PD	13-SEP-2001.			
PF	02-MAR-2001; 2001WO-JP01629.			
PR	07-MAR-2000; 2000JP-0159195.			
PA	(CHIB-) CHIBA PREFECTURE.			
FA	(HISM) HISAMITSU PHARM CO LTD.			
P1	Nakagawara A;			
DR	WPI: 2001-565584/63.			
TX	Nucleic acids originating in gene expressed in human neuroblastoma,			

PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents

XX Claim 1; Page 79; 2979pp; Japanese.

CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.

XX Sequence 910 BP; 326 A; 176 C; 156 G; 215 T; 37 other;

Query Match 16.5%; Score 722; DB 22; Length 910;
Best Local Similarity 93.1%; Pred. No. 2,4e-13;
Matches 787; Conservative 0; Mismatches 54; Indels 4; Gaps 4;

```
QY 3522 TTAATTCACAGGAGGCTCATCATCACACCTTAAGAGAGATTTCAGAAAAAAGTGG 3581
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 865 TTTATCCGNTGGGAGNTTATATACACNCCTTTAGGNGNGATTTTGGAAAAATGNN 806
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3582 CCAGATTTTCTTCTTCATCATTTTAAATGTGCGAGGCTTCAGTTTCTTACTCTT 3641
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 805 CCAGATTTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 747
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3642 ACCATGTGATTTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 3700
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 746 ACCNATGTGA-ATTTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 688
QY 3701 TTGACTTGTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 3760
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 TTGACTTGTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 628
QY 3761 TTTGGGTAGAGTTAGCAAAATTTAACCATTTGTGTTGCGCCATCCAGGAGGATCCCC 3820
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 TTTGGGTAGAGTTAGCAAAATTTAACCATTTGTGTTGCGCCATCCAGGAGGATCCCC 568
QY 3821 AGTTCTGACTTGAAGTAGAGTAGAGAGAAATCCAGAGGAGTCTATC-TGGCCAGATTAA 3879
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 AGTTCTGACTTGAAGTAGAGTAGAGAGAAATCCAGAGGAGTCTATC-TGGCCAGATTAA 508
QY 3880 GTAGATTTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 3939
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 GTAGATTTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 448
QY 3940 TTTAGGTTAAATCTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 3999
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 447 TTTAGGTTAAATCTCTTCTTCATCATTTTAAATGTGCGAGGNTGTCAG-TTTCNANTCTT 388
QY 4000 CAAGTTCCAGTCAAAATTTAGAGTGTGTTAGAGTGTGCGAGGNTGTCAG-TTTCNANTCTT 4059
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 CAAGTTCCAGTCAAAATTTAGAGTGTGTTAGAGTGTGCGAGGNTGTCAG-TTTCNANTCTT 328
QY 4060 TTTCCCTCATGAGTATACACACTGTGAATTCATCTTCACTGAAGGCCCTGCGAGTTCTC 4119
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 TTTCCCTCATGAGTATACACACTGTGAATTCATCTTCACTGAAGGCCCTGCGAGTTCTC 268
QY 4120 CTAAACATAGTGTGTTGTTTCTTAAACAAGTTTAAAGCTAGTTTAAATTAATTA 4179
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 CTAAACATAGTGTGTTGTTTCTTAAACAAGTTTAAAGCTAGTTTAAATTAATTA 208
QY 4180 AAAATGCTGTCTCTACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 4239
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 AAAATGCTGTCTCTACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 148
QY 4240 TTTGATTCATTAATTTTGAATTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCAT 4299
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 TTTGATTCATTAATTTTGAATTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCAT 88
QY 4300 CTGTAAGAAGTTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCAT 4359
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

DB 87 CTATTAAGAGTTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCATTTGATTCAT 28
QY 4360 AAAA 4364
DB 27 AAAA 23

RESULT 5
AAI94300/C
ID AAI94300 standard; cDNA: 893 BP.
AAI94300;
13-NOV-2001 (first entry)

Human neuroblastoma expressed polynucleotide SEQ ID NO 375.

Human neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

Homo sapiens.

WO200166719-A1.

13-SEP-2001.

02-MAR-2001; 2001WO-JP01629.

07-MAR-2000; 2000JP-0159195.

(CHIB-) CHIBA PREFECTURE.
(HISM) HISAMITSU PHARM CO LTD.

Nakagawara A;

WPI: 2001-565584/63.

Nucleic acids originating in gene expressed in human neuroblastoma,

useful as probe or primer in diagnosing prognosis of human

neuroblastoma, malignancy and susceptibility indicator or tumour marker

for anti-cancer agents

Claim 1; Page 318; 2979pp; Japanese.

The invention relates to novel genes (AAI93926-AAI97963) expressed in

human neuroblastoma. The nucleic acids are applicable as a probe or

primer in diagnosing the prognosis of human neuroblastoma, malignancy and

susceptibility indicators or tumour markers for anti-cancer agents. The

gene information for diagnosing prognosis is related to factors similar

to that for N-myc and TrkA genes.

Sequence 893 BP; 323 A; 171 C; 162 G; 215 T; 22 other;

Query Match 15.9%; Score 694.4; DB 22; Length 893;
Best Local Similarity 92.9%; Pred. No. 9.8e-132;
Matches 807; Conservative 0; Mismatches 54; Indels 8; Gaps 8;

```
QY 3498 CAAGATCTAAGAGAAATAGACTTAATTCACCTAGAGGCTCCATCTCAGACTTAAG 3557
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 882 CAGGATTTTGGGGGAATGGGACTTAATTCNNATAGGGCTTCATATTCACNCCTTAAG 823
QY 3558 GAGGAGATTTCTAGAAAAAAGTGGCCAGATTTTCTTGTCTCCATCATTTTAAATGAGC 3617
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 822 GNGGGGTTTNGAAAAAAGTGGCCAGATTTTCTTGTCTCCATCATTTTAAATGAGC 763
QY 3618 AGGCTGTCACTTTCTTACTCTTACCTAGTGAATTTCTTGTGAACGTGCCAAAAAG 3677
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 762 AGGTTG-TCACTTTCTTACTCTTACCTAGTGAATTTCTTGTGAACGTGCCAAAAAG 706
QY 3678 AAAAAGAGCCCAACAGAGCTCTGACTTTGTC-TTGAATCCCTCAGTTTC-TTCTTG 3735
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 705 GAAAAAGGCCCAACAGAGCTCTGACTTTGTC-TTGAATCCCTCAGTTTC-TTCTTG 646
QY 3736 ATTTCAGATGTGCGGTTCTTAATTTTGGGATGAGTGAAGCAATTAACATTTGTGT 3795
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

|||||
Db 645 ATTTCAGCATGTGTC-GGTTCCCTAATTTGGTATGATGAGCAAAATTTACCATTTGT 587
Qy 3796 TTGTGCCCTACCCAGGGGAGTCCCCAGTTCTCTACTGTGAAGTACTGAGAAATCCAC 3855
Db 586 TTGTGCCCTACCCAGGGGAGTCCCCAGTTCTCTACTGTGAAGTACTGAGAAATCCAC 527
Qy 3856 -GAGTGTCTATC-TGGCCAGATTTAAGTATCTATTTCTCTGTTCTCCCTCCCTG 3913
Db 526 GAGAGTCTATCTTCCCAATTTAAGTATCTATTTCTCTGTTCTCCCTCCCTG 467
Qy 3914 AGAGCTCTATTTATTTATGTCCTCTCTGAGTTAATCTCTGATTTGACTTTGTT 3973
Db 466 AGGACCTCTATTTATTTATGTCCTCTCTGAGTTAATCTCTGATTTGACTTTGTT 407
Qy 3974 GAGAGGAGTGTGAGCATGATTTAGCAAGTTCGCAAGTCCAAATTTAGAGTGTAGA 4033
Db 406 GAGAGGAGTGTGAGCATGATTTAGCAAGTTCGCAAGTTCGCAAGTTCGATTTAGA 347
Qy 4034 GTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGAGTACAACTGTGAATTCAT 4093
Db 346 GTGTGGGGGAAAAATTAGTCTTATTTTCCCTACATGGAGTACAACTGTGAATTCAT 287
Qy 4094 CTTCACATGAAGGCCCTGCGAGTCTCTCTAAACATAGTTGTTGTTTCTTTAACAAG 4153
Db 286 CTTCACATGAAGGCCCTGCGAGTCTCTCTAAACATAGTTGTTGTTTCTTTAACAAG 227
Qy 4154 TTTTACGTTAGTTAATTAATTTAAAAAATGCTTCTGCTGCTGCTGCTGCTGCTGCT 4213
Db 226 TTTTACGTTAGTTAATTAATTTAAAAAATGCTTCTGCTGCTGCTGCTGCTGCTGCT 167
Qy 4214 ATGCCATTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4273
Db 166 ATGCCATTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 107
Qy 4274 TGTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4333
Db 106 TGTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 47
Qy 4334 GGCTGTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 4362
Db 46 GGCTGTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAA 18

```

RESULT 6
AA194299 standard; cdna; 903 BP.

AA194299;
13-NOV-2001 (first entry)

Human neuroblastoma expressed polynucleotide SEQ ID NO 374.
Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

Homo sapiens.
WO20016719-A1.
13-SEP-2001.
02-MAR-2001; 2001WO-JP01629.
07-MAR-2000; 2000JP-0159195.
(CHIB-) CHIBA PREFECTURE.
(HISM) HISAMITSU PHARM CO LTD.
Nakagawara A;
WPI: 2001-565584/63.

```

PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents
PS Claim 1; Page 317; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AA193926-AA197963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
SQ Sequence 903 BP; 257 A; 169 C; 197 G; 250 T; 30 other;

Query Match 15.3%; Score 670; DB 22; Length 903;
Best Local Similarity 93.2%; Pred. No. 9.2e-127;
Matches 780; Conservative 0; Mismatches 48; Indels 9; Gaps 8;

Qy 1830 CAGAAATTTGAGCTACCTGCTGTTGATGAGCGGTAGCCGATGCGAATT-GGAGATCCAGAA 1888
Db 69 CAGAAATTTGAGCTACCTGCTGTTGATGAGCGGTAGCCGATGCGAATTGCGAATCCAGAA 128
Qy 1889 GAAGCAAAACCTCCACCGGAGCTGTAGGAATAGCTGTGTGCGCAAGAACCTGTCTCA 1948
Db 129 GAAGCAAAACCTCCACCGGAGCTGTAGGAATAGCTGTGTGCGCAAGAACCTGTCTCA 188
Qy 1949 GATAGTTGTAGCATGCGCATCCCGAGAGTGGAGACCGTGTATGTGACCTTTGCT 2008
Db 189 GATAGTTGTAGCATGCGCATCCCGAGAGTGGAGACCGTGTATGTGACCTTTGCT 248
Qy 2009 CACATATGTATACATGCGTGTATATACCTTTTCACTTCTGACTTTGTTTCATTA 2068
Db 249 CACATATGTATACATGCGTGTATATACCTTTTCACTTCTGACTTTGTTTCATTA 308
Qy 2069 CTCTGATTTACAAAAAATCTTTTCACTTCTGCTTATGTGATTAAGAGGATGATGGGA 2128
Db 309 CTCTGATTTACAAAAAATCTTTTCACTTCTGCTTATGTGATTAAGAGGATGATGGGA 368
Qy 2129 TTTCTTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATATAGATGGC-AGATTTC 2187
Db 369 TTTCTTTTCCCTTTTGGGAAATGGGCTCTCAAGCTAAAGCTATATAGATGGCAGATTTC 428
Qy 2188 AGAAGTTTCAAGGGGCTGTTTCTATACATTTGGCTATGTTAAAGGGGTAAAGGGCTCTC 2247
Db 429 AGAAGTTTCAAGGGGCTGTTTCTATACATTTGGCTATGTTAAAGGGGTAAAGGGCTCTC 488
Qy 2248 TTTCAATAGCATGTGGAAGATGAAGACAGCCCTCTCTTACCTGCTGCTGCTGCTGCTG 2307
Db 489 TTTCAATAGCATGTGGAAGATGAAGACAGCCCTCTCTTACCTGCTGCTGCTGCTGCTG 548
Qy 2308 TCTTCTACCCCTGTGACACCTCTCTTATATAGTGGAT-AGTGAATTTTAAACCTTAATA 2366
Db 549 TCTTCTACCCCTGTGACACCTCTCTTATATAGTGGATTAAGTGAATTTTAAACCTTAATA 608
Qy 2367 AAACAAACACCTCCACCATGAGCTTTAGAGC-AGAAGAGGAATGAC-AAAGCAAGCAT 2424
Db 609 AAACAAACACCTCCACCATGAGCTTTAGAGCAAGAGGAATGACAAAGTGAACGAT 668
Qy 2425 GAAGCAAGCATCTTCA--GAGTGAAGAAAGCATCGAGCTGTTGATATAGTGTCT 2482
Db 669 GAAGCAAGCATCTTCAAGAGTGAAGAAAGCATCGAGCTGTTGATATAGTGTCT 728
Qy 2483 GAAAGATAGTGTCTTCAAGTGAAGAACTGTGATACATCTGAGGAGGATGTT 2542
Db 729 GAAAGATAGTGTCTTCAAGTGAAGAACTGTGATACATCTGAGGAGGATGTT 786
Qy 2543 GGCTGTGATTAATTTTCACTTAATGATTAACATTTCTTACTGCTCAAAAACCAAAAT 2602
Db 787 GGCTGTGATTAATTTTCACTTAATGATTAACATTTCTTAAACGNCMAAAACCAAAATC 846
Qy 2603 CTTTGGAAAAGAAAGTGGGATGTTAGTTTCAAGAACATGCTACAGCTGAACAAA 2659

```


AC AAC76874;
 XX 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2429 polynucleotide sequence SEQ ID NO:4857.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparinsonian; nootropic; hepatoprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antihemetic; antihypoid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disease; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB42665.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4044-4045; 5507pp; English.
 XX
 XX AAC74446 to AAC7606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparinsonian; antiparinsonian; nootropic; hepatoprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antihemetic;
 CC antinaemic; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancer;
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, coagulation,
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 2887 BP; 796 A; 659 C; 691 G; 739 T; 2 other;
 Query Match 8.7%; Score 380.2; DB 21; Length 2887;
 Best Local Similarity 95.6%; Pred. No. 1,4e-67;

Matches 391; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 1045 GGGGAGGTGGAAGGAGGACATTAAGAAATCCCATTTGGAGTACAGAAAGAGACAC 1104
 DB 2493 GGAGAGGGGTGGAGGGGTTAGACTTCTACCCCATTTGGAGTACAGAAAGAGACTC 2434
 QY 1105 CTGTTAAAAAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAAAACCTCTAGCACTTC 1164
 DB 2433 CTGTTAAAAAGCTGGCTCTGTAATTTTCAAAAGTCAAAACAAAACCTCTAGCACTTC 2374
 QY 1165 CTATTAAAGAGAACCTCTGGTCTCTATCTCAAACTGTTTAAACGTGAATTAGAGA 1224
 DB 2373 CTATTAAAGAGAACCTCTGGTCTCTATCTCAAACTGTTTAAACGTGAATTAGAGA 2314
 QY 1225 GCCAAGAAACCCAGAAAGCCCGGTCTCACTGAGACACCCAGCAAGCTCTCCACTC 1284
 DB 2313 GCCAAGAAACCCAGAAAGCCCGGTCTCACTGAGACACCCAGCAAGCTCTCCACTC 2254
 QY 1285 CCCAAAGGAGCCAGCCAGCCATCCCAAGAGAAACCTTCTCACTGAGATGAAGATT 1344
 DB 2253 CCCAAAGGAGCCAGCCAGCCATCCCAAGAGAAACCTTCTCACTGAGATGAAGATT 2194
 QY 1345 TGCCGTACCTTTCCACACAGAAATGTTGTTGGACACGCTCCGCCATCAC 1404
 DB 2193 TGCCGTACCTTTCCACACAGAAATGTTGTTGGACACGCTCCGCCATCAC 2134
 QY 1405 CGTTACCATTTACGGAATCCTCTCCAAAGAGAGAGAGACTGTAGCAAG 1453
 DB 2133 CGTTACCATTTACGGAATCCTCTCCAAAGAGAGAGAGACTGTAGCAAG 2085
 RESULT 12
 AAC25273
 ID AAC25273 standard; cDNA; 317 BP.
 XX
 AC AAC25273;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 29348.
 DE
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 OS
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 9905-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Maline Edwards J, Duclert A, Giordano J;
 XX
 DR WPI; 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 29348; 71pp + CD-ROM; English.
 XX
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences

P1 Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

[illegible]

4145 TTACCAAGTTTAAAGCTAGTGTAAATAAATTAAA--AAAAATTGCTTGTCTGTCTACTTC 4202

Db 197 TTAACAAGTTTAAGCTAGTGTAAATAAATAAGAGAAGAGCTGTGCTCCACCTC 138
QY 4203 AGCTTGTGTTTAAAGCCATTTCATATGTTGCTGTGTTGTAATATATACTTTGATAC 4262
Db 137 AGCTTGTGTTTAAAGCCATTTCATATGTTGCTGTGTTGTAATATATACTTTGATAC 79
QY 4263 CATTTCTGATGCTAAATATGTTGCTGTGTAATATATCTTAATAAGAGTTCAATTTGTAA 4322
Db 78 CATTTCTGATGCTAAAGCTGCTGTGTAATATATCTTAATAAGAGTTCAATTTGTAA 19
QY 4323 TAAACTATGTGCTGCTGT 4340
Db 18 TAAACTATGTGCTGCTGT 1

RESULT 14

AAH36707 standard; cDNA; 196 BP.

AAH36707;

03-SEP-2001 (first entry)

Human colon cancer antigen encoding cDNA SEQ ID NO:3789.

Human: colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma; ss.

Homo sapiens.

WO200122920-A2.

05-APR-2001.

28-SEP-2000; 2000MO-US26524.

29-SEP-1999; 99US-0157137.

03-NOV-1999; 99US-0163280.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA.

WPI; 2001-235357/24.

P-PSDB; AAG77300.

Claim 1; Page 5645-5647; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated P, by inserting the nucleic acids into a host cell and culturing the cell, to express the proteins. N and P may be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 196 BP; 39 A; 42 C; 42 G; 62 T; 11 other;

Query Match 4.0%; Score 176.2; DB 22; Length 196;
Best Local Similarity 93.2%; Pred. No. 2e-26;
Matches 178; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3810 GGGGACTCCCACTTCTGACTGAGTAGACGAGAAAGAACCCAGAGGTGCTATCTCG 3869
Db 5 GGNNTTCCCACTTCTGACTGAGTAGACGAGAAAGAACCCAGAGGTGCTATCTCG 64
QY 3870 CCAGATTTAAGTAGATTCTATTTCCCTGCTCCCTCCGAGACCTGTTATTTTA 3929
Db 65 CCAGATTTAAGTAGATTCTATTTCCCTGCTCCCTCCGAGACCTGTTATTTTA 124
QY 3930 TTGTCCTCCCTCTCTAGATTAAATTCCTTTGATTGAGTGAAGAGAGTTGAC 3989
Db 125 TTGTCCTCCCTCTCTAGATTAAATTCCTTTGATTGAGTGAAGAGAGTTGAC 184
QY 3990 ACTAGATTAGC 4000
Db 185 ACTAGATTAGC 195

RESULT 15

ABV56446 standard; cDNA; 211 BP.

ABV56446;

17-SEP-2002 (first entry)

Human prostate expression marker cDNA 56437.

Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

Homo sapiens.

WO200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001MO-US05171.

17-FEB-2000; 2000US-183319P.

16-MAR-2000; 2000US-189862P.

25-MAY-2000; 2000US-207454P.

09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.

13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 10890; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (1) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate, cell carcinogenic potential of a compound;

(g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

SO Sequence 211 BP; 89 A; 37 C; 29 G; 55 T; 1 other;

Query Match 3.1%; Score 134.8; DB 23; Length 211;

Best Local Similarity 98.6%; Pred. No. 5.6e-18;

Matches 136; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4235 CTGTGTGTAATTCATTAACCTTTGATACCATTTCTGATGTTGTA 4294

Db 48 CGGGGTGTGTAATTCATTAACCTTTGATGTTGTA 107

QY 4295 AATATCTTATPAAAGATTCAATGTAATAAATGCTGTA 4354

Db 108 AATATCTTATPAAAGATTCAATGTAATAAATGCTGTA 167

QY 4355 AAAAAAAAAAAAAAAAAA 4372

Db 168 AAAAAAAAAAAAAAAAAA 185

Search completed: July 14, 2003, 12:04:01
Job time : 871 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 11:47:11 ; Search time 43 seconds

(without alignments)
1372.710 Million cell updates/sec

Title: US-10-054-935-2

Perfect score: 3338

Sequence: 1 MTKRSAYFKAAAPAGGNPE.....RSRCRLFIQKQTPHRCRK 614

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211	6.5	955	2	male-specific leth
2	202	6.2	854	2	neurofilament trip
3	200	6.2	3938	2	neurofilament trip
4	199	6.1	1020	1	neurofilament trip
5	197.5	6.1	1072	1	neurofilament trip
6	197.5	6.1	3942	2	neurofilament trip
7	192	5.9	3534	2	neurofilament trip
8	190	5.9	1087	1	neurofilament trip
9	190	5.9	1098	2	neurofilament trip
10	189.5	5.9	606	2	neurofilament trip
11	189	5.8	539	2	neurofilament trip
12	188	5.8	2715	2	neurofilament trip
13	186.5	5.8	460	2	neurofilament trip
14	184.5	5.7	1520	2	neurofilament trip
15	183.5	5.7	1520	2	neurofilament trip
16	182	5.6	1216	2	neurofilament trip
17	181.5	5.6	2187	2	neurofilament trip
18	178	5.5	1585	2	neurofilament trip
19	176.5	5.5	7962	2	neurofilament trip
20	173	5.3	351	1	neurofilament trip
21	173	5.3	1207	2	neurofilament trip
22	173	5.3	2022	2	neurofilament trip
23	172.5	5.3	707	2	neurofilament trip
24	172.5	5.3	1188	2	neurofilament trip
25	172	5.3	1151	2	neurofilament trip
26	172	5.3	1870	2	neurofilament trip
27	172	5.3	1870	2	neurofilament trip
28	171.5	5.3	3968	2	neurofilament trip
29	171	5.3	1285	2	neurofilament trip

30	171	5.3	1647	2	SNF2beta protein -
31	170.5	5.3	418	2	hypothetical prote
32	170	5.3	834	2	hypothetical prote
33	169.5	5.2	858	2	neurofilament trip
34	169.5	5.2	891	2	neurofilament trip
35	169	5.2	268	2	probable prolina-r
36	169	5.2	1613	2	homeotic protein C
37	169	5.2	2342	2	protein BRG1 - hum
38	168.5	5.2	3530	2	hypothetical prote
39	167	5.2	398	2	unconventional myo
40	167	5.2	657	2	hypothetical prote
41	167	5.2	1320	2	probable Sfl6 prot
42	167	5.2	2142	2	TCOF1 protein - mo
43	166.5	5.1	6642	2	MHC class III hist
44	166	5.1	753	2	protein UNC-89 - C
45	166	5.1	1634	2	Op protein - Kenne
					hypothetical prote

ALIGNMENTS

```
RESULT 1
S52959
male-specific lethal-1 protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
C:Accession: S52959; S65350
R:Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.
Genetics 134, 545-557, 1993
A>Title: The male-specific lethal-1 (msl-1) gene of Drosophila melanogaster encodes
A:Reference number: S52959; MUID:93314941; PMID:8325488
A:Accession: S52959
A:Molecule type: DNA
A:Residues: 1-955 <PAL>
A:Cross-references: EMBL:L14582
R:Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.
submitted to the EMBL Data Library, April 1993
A:Reference number: S65350
A:Accession: S65350
A:Molecule type: DNA
A:Residues: 1-407, 'S', 409-585, 'M', 587-955 <PAL>
A:Cross-references: EMBL:L14582
C:Genetics:
A:Gene: msl-1
A:Cross-references: FlyBase:FBgn0005617
A:Introns: 329/3

Query Match 6.5%; Score 211; DB 2; Length 955;
Best Local Similarity 20.3%; Pred. No. 0.0061;
Matches 96; Conservative 91; Mismatches 171; Indels 114; Gaps 18;

101 PPLATATAGTAAEGKMKRSPLOGGSGSSAQAACIKITLL-----LQDLIE 234
485 PKHLKAAVAPKVTKTSRSTLPKA-----NTADIDAPAKQYIAHNGSTQTDTDK 537
235 QOOOOLQAKKEI-----ELKSERDT-LIARLERRMQVLKDNKERRKL 282
538 TORLOVKTRQYEMHMDMTGSSAPSDIKQKQVNDSPSPERTKTSKSIIVNDKTTSET 597
283 FQGYETREERET--ELSEKIKLECOPELSETSTLP-----PKFSGRGSK 327
598 SQSPQDELIDVEYVRKLAHLKELLSQSHSQVTLKIRERVAATNLVPP----- 649
328 GHKRSPPGSTERKTPVKLAPEKSVKTKPKH---SPIKEPQGSISEYCKRELASQ 384
650 ----SAPVST---TTPAPPTPTPTGSPPOHAVTSMQDEISAASKSAAROI--- 699
385 EPPEKPRSSVTPPLPLSTPQK-----GSPTHKEKAFSEIED-----LPLSTT--- 429
700 ATPLTPQSSSVSSTSTIRKTLNNGSPHTYSKATARSGKQSRRTATFPYSTTWEDQ 759
430 -----ENYLGRHWPSPPLPLRESSPKKEETVAVARCLMPSSVAGETSVLAVPSWRDSV 483
```

Db 760 EFHCDNEFL-----EEA-----DELLADNPSLEIPKRDV 792
Qy 484 EPLRDPNPSDLLENLDDSVFSKRAKLELDEKRRKMDIQIRIORIOLQRLMYKKG 543
Db 793 PPSSDKIDTEL---LSQATERRRHKQYKVDKDEVCARDARKMKQIRLEQJLMMKRNDEV 849
Qy 544 IQESEP-EVTSFEPPDDESIMTTPPLPVVAFGRPLPKLPQ--NFELPWLD 593
Db 850 LVALDPLRASTFYPPLPEDIEAIFVNEVTVQAFGENVNMARDFGVPMWD 901

RESULT 2

S02003
neurofilament triplet H protein - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 13-Aug-1999
C:Accession: S02003
R:Breen, K.C.; Robinson, P.A.; Wion, D.; Anderton, B.H.
FEBS Lett. 241, 213-218, 1988
A:Title: Partial sequence of the rat heavy neurofilament polypeptide (NF-H). Identification
A:Reference number: S02003; MUID:89065087; PMID:3143606
A:Accession: S02003
A:Molecule type: mRNA
A:Residues: 1-854 <BRED>
A:Cross-references: EMBL:X13804; NID:957828; PIDN:CAA32038.1; PID:957829
A:Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil

Query Match 6.2%; Score 202; DB 2; Length 854;
Best Local Similarity 21.5%; Pred. No. 0.015;
Matches 121; Conservative 92; Mismatches 252; Indels 98; Gaps 22;

Qy 11 AADAGNPNRQRLDYERAAALGGPEDEPGAAEHFL---LPRHRLKEPGRPLASSQGGSP 67
Db 315 AKSPA-----EVKSPAIEYKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPA 366
Qy 68 APSNAGCGGKGRILLNAGA-----APGQDESMGSGVPLPCPPATQAGIGGE-PA 119
Db 367 AEVKSPEVAKS-----PAEAKSPAIVKSPGEAKSPAIAKSPAIAKSPAIAKSPAIAKSPA 421
Qy 120 AAGAGCSPRPYQAVLPIOTGSLVAAKEPTPMAGDRGGAASPAATASDPAGPPPLPG 179
Db 422 EVKSPVIVKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPA 471
Qy 180 PPPLAPATAGTTLAASGRKMSMKRSPGGGGGGAASQAACLQILLDLLEQQQQQ 239
Db 472 KSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPAIAKSPA 528
Qy 240 LQAKKEIEELKSRDILLARIEMERRMQLVKDNEKRRKLEFGYETETRETELSEK 299
Db 529 VKEGAKSLAEKKSPEKA-----KSPVKEIKPPAEVKSPEKAKSPMKR 572
Qy 300 IKLECOPELSETSOTLPP--KPFSGRSGKHKRSPFGSTER-KTPVKKLAPFSVKT 356
Db 573 AK---SEKAKKIDLVKSPKAP---PAKEEAKRPAIRSPVOYKSPAIAKSPAIAKSPA 625
Qy 357 KTRKHSIKPEPCGSLSETVCK---RELRSQETPEKRRSV-----DTPRLSTP--Q 404
Db 626 RTEVAAKKEEVKSPVEEVAKKEPPKVEEEKTPATKTEVKEKSKDEAKQKPAEE 685
Qy 405 KGPSTH--PKKASSELEDLPYLSTEMYLCRMHQPPSPPLPLR--ESSPKKEEVAVACLM 462
Db 686 KEPLTEKPKDSPEAKKEE-----AKKKAAAPDEETPAKLVGEEKPKPEKADAAKE 740
Qy 463 PSSVA-----GETSVLAVPSMRDHSVE-----PLRDPNPSDLLENLDDSVFSKR 506
Db 741 PSNPSEKPKKEEVPAPEKPKDKEKTEKSKKREKPKMEPRPKRRTAFAFKSLANPR 800
Qy 507 HAKLELDEKRRKMDIQIRIREQR 529
Db 801 QKRLKSPLAQTKKTASQRRPQR 823

RESULT 3

T42761

Bassoon protein - rat

N:Alternate names: brain-specific synapse-associated protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42761

R:Dieck, S.; Sammarti-Villa, L.; Langeneese, K.; Richter, K.; Kindler, S.; Soyke, A.; W
J. Cell Biol. 142, 499-509, 1998A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively local
A:Reference number: 222249; MUID:96345363; PMID:9679147

A:Accession: T42761

A:Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: mRNA

A:Residues: 1-3938 <DIE>

A:Cross-references: EMBL:Y16563; NID:93413503; PIDN:CAA76287.1; PID:93413504

A:Experimental source: strain Sprague Dawley; brain

A:Function:

A:Description: may be involved in cytomatrix organization at the site of neurotransmit

A:Note: component of the presynaptic cytoskeleton

C:Keywords: coiled coil; zinc finger

Query Match 6.2%; Score 200; DB 2; Length 3938;
Best Local Similarity 20.4%; Pred. No. 0.089;
Matches 162; Conservative 56; Mismatches 250; Indels 328; Gaps 32;

Qy 3 MRSVAFKAAAPAG---GNPEQRLDYERAAALGGPEDE-----PGAAEHFLPRHRK 51
Db 239 LHPSPALPSAIPKAPQPKPEQERSRSPATQSPRAEAARATSVGPQATAPPEVGR 298
Qy 52 LK-----EPGPPLASSOGGSPADSPACG-----GKGRGILL- 83
Db 299 VSPQPLSTRPSTAEPRPPAGEAQKSGATTVPGLAAEOTGGTGLKGLGLASLITQA 358
Qy 84 -----PAAAPGQGESEMG-----GSVLPCCPPATKQ----- 111
Db 359 STLMSVQPEADTQGPSPSKGPKYIVSDASKENGPRPPSGGPGPTPAKTEPGRIG 418
Qy 112 -----AGIGGEP-----AAAGAGCSPR--PKYQAVLPIOTGSLVAAK----- 147
Db 419 PGSGPGLATGTGTPSPKHKRQADQAAKAAAPKTKPKRAACPLQALNANGSGPAN 478
Qy 148 -----EPFPMAGDK-----GGAASPAATASDPAGPPPL 175
Db 479 YNTACKLRYCTLCGFNPPLHVEKTEWMLCLNCOTKRLLEGSLGEPA-----PLPL 530
Qy 176 PLPGPPLATPATAGTTLAASGRKMSMKRSPGGGGGSG-----ASSQAACLQILLDL 230
Db 531 PTFQDEP-----AGVQRAAGASPLKQKGPQGGPSSGLPPKASFOAKKASP-----QA 580
Qy 231 DLLEQQQQQLOAKKEIEELKSRDILLARIEMERRMQLVKDNEKRRKLEFGYETEE 290
Db 581 AKASPOAKPLRASP-----PS 596
Qy 291 REETELSEK---IKLECOPELSETSOTLPPKPPSCGSGKHKRSPFGSTERKTPYKTL 347
Db 597 KTSSEAEKKTGTPVKAEP-----VKKP-----PPETAVPPTPKAKSGVKRT 639
Qy 348 APERSKVKTKTPKHSPIKEEP--CGSLSETVCKRELRSQETPEKRRSVDPPLRSTPQK 406
Db 640 DP-----ATPVVKKPPEAPKSGEAEPPK--TSQDLSRSPQSLSDI-----GVSSDG 686
Qy 407 PSTHPRK--KAFSSELEDLPYLSTEMYLCRMHQPPSPPLPLR--SSPKKEETVARCLMP 463
Db 687 VSSQSEITGVVQOEVLQDSAGVT-----GRRPSPSELHKVSGSMKPSLEAQAQAVAP 739
Qy 464 S---SVAGETSVLAVPSMRDHSVEPL-----RDPNPSDLLEN 497
Db 740 SGEMSKPPSGSAVEDQKRRPHSLSIMPEAFSDDELGLDLEEDDSLAMGQRQOQTAES 799
Qy 498 LDDSVFSKRAKLE-----LDEKRRKMDIQIRIREQR----- 533
Db 800 SDFGSGQLRHVDYEDSSGGLSPLPPQPARADMTDEEFMRKQITLESKAEEDNLEDDTA 859

QY 534 LQRMVKKKGIQSEPEVTSFEPEDDVESLMTPLFVAFGRPLKTPONFPLPMD 593
 DB VSGRGLAKHGAOK-----ASARPPRESSQESVALP---- 889
 QY 594 ERSRCRLERIQKKQTPH 609
 DB 890 -----KRLRPH 895

RESULT 4

OPH0H
 neurofilament triplet H protein - human
 N:Alternate names: neurofilament protein, 112K
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
 C:Accession: S00979
 R:Lees, J.F.; Smeidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
 EMBO J. 7, 1947-1955, 1988
 A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
 A:Reference number: S00979; MUID:88328981; PMID:3138108
 A:Accession: S00979

A:Molecule type: DNA
 A:Residues: 1-1020 <LEE>
 A:Cross-references: EMBL:X15306; NID:g935028; PIDN:CAA33366.1; PID:g1841430
 A:Note: It is uncertain whether Met-1 or Met-2 is the initiator
 C:Genetics:
 A:Gene: GDB:NEFH
 A:Cross-references: GDB:120225; OMIM:162230
 A:Map position: 22q12.1-22q13.1
 A:Introns: 295/1; 361/3; 403/2
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
 F:1-100/Domain: amino-terminal <NTP>
 F:101-410/Domain: rod #status predicted <ROD>
 F:411-1020/Domain: carboxyl-terminal <CTD>
 F:503-826/Region: 14-residue repeats
 F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
 (covalent) #status predicted
 F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 6.1%; Score 199; DB 1; Length 1020;
 Best Local Similarity 23.6%; Pred. No. 0.024;
 Matches 119; Conservative 70; Mismatches 230; Indels 86; Gaps 22;

QY 90 GQOESMGSVPLPCPPATKQAGIGEPAAAGCSPRKYQAVLP-IGSLVAAKE 148
 DB 489 GEEEAEGEGEETKSP-----AEEAAS-----PEKEAKSPVKEAKSPAEKAS 532
 QY 149 PTPMAGDKGAASPAAT-----ASDPA-----GPPPLPLPGPPPLPATAGTLAASGR 198
 DB 533 P-----EKKEAKSPAEVSPAEKAPAEKASPEKASPEKASPEKASPEKAS 587
 QY 199 WSMKRSPLGGGGGSGAQAACLKQLLLDLLEQQOQOQLOAKEKEIEELSERDYL 258
 DB 588 AKKEAKSPAEKSPAEKASPVK-----EKAASPAEKSPVKEAKSP----- 629
 QY 259 ARIEMERMLVKKDNKEKRLKLFQGYETERETLESEKILCEQPELSTSTQTLPPK 318
 DB 630 AEVKSPEK-----AASPTKEAKSPAEKASPEKASPEKASPVKAEKASPEKASPV 684
 QY 319 PFSCGSGKHGRKSPFESTERTKTPVKKLAPFSSVTKTKTPKHSPIKE-PCGSLSSEVC 377
 DB 685 K---AEAKSPAEKSPV-KEEAKSPAEKASPVKEAKSPAEKASPVKEAKSPAEKASPV 740
 QY 378 KRELRSQE---TPKPRSSVDTPPLSLTPQKGPSTHP---KEKAFSEIEDLPLSTTE 430
 DB 741 KEKASPEKASPEKAKTLDVKSPAEKAPAEKASPADKPEKASPVKEKASPVKEKASPV 800
 QY 431 MYLCRMHQPSPPLRE--SSPKKE---IVARCLMSSVAGETSVLAIPSM---RDHS 482
 DB 801 SPLKADAKAPKEITPKKEKSPVKEEKPOEVKKEPKKAEKADAPTKTEKKDSK 860

QY 483 VE--PLRPNNSDLLENIDSVFSKRHAKLEDEKRRKRWIDQRI--REORTIQLQRM 538
 DB 861 KEAKKRAKPAKPKVEKEKEPAVEKKEKVE--AKKEADKKKVPTEKEKAPAVEV-- 916
 QY 539 YKKKGIOSEPEVTSFEPEDDVES 563
 DB 917 -KEDAKPKTEVAK--KEPDAKA 938

RESULT 5

A37221
 neurofilament triplet H protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
 C:Accession: A37221 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
 R:Chin, S.S.M.; Liem, R.K.H.
 J. Neurosci. 10, 3714-3726, 1990
 A:Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with
 A:Reference number: A37221; MUID:91038277; PMID:2230956
 A:Accession: A37221

A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1072 <CHI>
 A:Cross-references: GB:AF031879; NID:g2642597; PIDN:AAB87068.1; PID:g2642598
 R:Robinson, P.A.; Wion, D.; Anderton, B.H.
 FEBS Lett. 209, 203-205, 1986
 A:Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
 A:Reference number: A25649; MUID:87080760; PMID:2878828

A:Accession: A25649
 A:Molecule type: mRNA
 A:Residues: 230-318; 472-542 <ROB>
 A:Cross-references: GB:M37227
 R:Dautigny, A.; Pham-Dinh, D.; Rousset, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.
 Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988
 A:Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ
 A:Reference number: A30796; MUID:88309090; PMID:2457365
 A:Accession: A30796

A:Molecule type: mRNA
 A:Residues: 266-421, 'V', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', '7
 A:Cross-references: GB:M21964; NID:g205685; PIDN:AAA41695.1; PID:g205686
 R:Leberburg, I.; Spinner, N.; Snyder, S.; Anderson, J.; Goldhaber, D.; Smulowitz, M.
 Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989
 A:Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament pepti
 A:Accession: A32757; MUID:89184647; PMID:2928342

A:Reference number: A32757; MUID:89184647; PMID:2928342
 A:Accession: A32757
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIE>
 A:Cross-references: GB:J04517; NID:g205679; PIDN:AAA41692.1; PID:g205680
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.1%; Score 197.5; DB 1; Length 1072;
 Best Local Similarity 22.1%; Pred. No. 0.03;
 Matches 112; Conservative 82; Mismatches 235; Indels 77; Gaps 20;

QY 11 AAAPGNGPEQRLDYERAAALGGPEDEGAFAHLPNHRK---LKEPGPLASSQGS 66
 DB 548 AEAKSPAEKSPAEKSPATYKSPAEKSPAEK--SPAENVSPATYKSPGEAKSPAEKAS 606
 QY 67 PAPSPAGGGGGRGLLTPAGA-----APGQOESMGGSVPLPCPPATKQAGIGGE-P 118
 DB 607 PAEVKSPAEKAS-----PAEAKSPASVSPGEAKSPAEKSPAEKSPATYKSPAEKAS 661
 QY 119 AAAGAGCSPRKYQAVLP-IGTSLVAAKKEPTPMAGDKGAASPAATASDPAGPPPLP 178
 DB 662 AEVKSPTVTKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 719
 QY 179 GPPPLPATPATAGTLAASGRKMSKRSPLGGGGGSGAQAACLKQLLLDLLEQQOQ 238
 DB 720 SPAEAKPPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 768

QY 239 QLAKEIEELKSERDITLARIEMRRMOLVKNKNEKHKILPGYEETEETELSE 298
DB 769 PVKGAISLAEAKSPEKA-----KSPVKEIKPPAAVKSPEKAKSPMKE 812
QY 299 KIKLECOPELSETQTLIPKPFSCGRSGKHKRKSPPGSTER-TTPYKTLAPESKATK 357
DB 813 EAK---SPEKAKTLDVKSPEAKTPAK--EAKKRRADIRSPQVAKSPAKKEEKSEKETR 867
QY 358 TPKHSPKEPCGSLSETVCK---RELRSQETPEKPRSSV-----DTPPRLSTP---OK 405
DB 868 TEKVAPEKEEKSPVEEVKAKKEPKKVEEETPTAPTEVEKSKKDAKPAKPAKDEK 927
QY 406 GPSTH-PKEKAFSEIEDLPLSTTEMYLCRMHQPSPPLPLR-ESSPKKEETVAKCLMP 463
DB 928 EPLTEPKDPSGEAKKEE-----AKEKKAAPPEETPAKLGVEKPAKPEKADAKAKAP 982
QY 464 SSVA-----GETSLAVPSMRHSVE 484
DB 983 SKPSEKPKKEEVPAAPEKKDTKEE 1008

RESULT 6

T42730

Bassoon protein - mouse

C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Dieck, S.; Sanmarti-Villa, L.; Langnese, K.; Richter, K.; Kindler, S.; Soyke, A.; Mex, J.; Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
A:Reference number: Z22249; MUID:98345363; PMID:9679147
A:Accession: T42730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3942 <DIE>
A:References: EMBL:Y17034; NID:93413809; PIDN:CAA76598.1; PID:93413810
C:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9P1
A:Insertions: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
A:Note: bassoon
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmitter
A:Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

Query Match 6.1%; Score 197.5; DB 2; Length 3942;
Best Local Similarity 22.5%; Pred. No. 0.12; Mismatches 173; Indels 241; Gaps 32;
Matches 138; Conservative 62;

QY 32 GGPEDFGAAEAHFLPRHRLKEPGLASSQ-----GSPAP----- 69
DB 2125 GGP-----LVQYQPOH-----GPLSAPOGLAPRLSGILGNPTREGPSGNLAQY 2172
QY 70 SPACGGKRGILLPA-----GAAPQ-----OES----- 95
DB 2173 GPAASQATAVRQQLPSTAVYAAAGMITYSTINPIAATLPTTQPAVLRPMVNGMYRP 2232
QY 96 --WCG--SVPL-----PCP-----PATKQAGIGGEPA 119
DB 2233 YVSGGVAVPLTSLTRVPMIARVPLGAPAGYRPAIRPFIASSVPAEGVYL-GKPA 2221
QY 120 --AAGACSPPKYQAVLPQTGSLVAAAEPTPWAGDGGASPAATASDPAPPLPL 177
DB 2292 TKASGAGPPRPPELPAGV-----AREEPF-----STTAPAVIKAPVAPGPA 2335
QY 178 PGPPPLPATAGTLASSEGKSMKSPGLGGGSGASASQACILKILLLDILIQOQ 237
DB 2336 PAPP-----GKPPAGEA-----AAGSGSVLSRPASEKE-----BASQ 2369
QY 238 QOQAKKEIEELKSERDITLARIEMRRMOLVKNKNEKHKILPGYEETEETELSE 297
DB 2370 EDROKQOE-QILOLEKE-----KVELEKLOLRLOEELERBRVEL-----QRHNEEQLL 2419

QY 298 EKIKLECOPELSETQTLIPKPFSCGRSGKHKRKSPPGSTERTPVKTLAPESKATK 350
DB 2420 -----VQRELOEL-QTI-----KQHVLDQOQOEERQAFALQREQLAQORLOLE 2461
QY 351 FSKVKTTPKHPSPKEPCGSLSETVCKRELRSQETPEKRRSSVDTPPRLSTQKGPSTH 410
DB 2462 -----QIQO-----LQOOLQLOLEQOKORAKAPPATCAEAPSKGPPA 2499
QY 411 PKEKAFSEIEDLPLSTTEMYLCRMHQPSPPLPLRESPPKKEETVAKCLMPSSVAGET 470
DB 2500 ATELAQNGQY--WPLTHAFLVAGTEGGQ-----REPLYLHGLDSS-ASDM 2546
QY 471 SVLAVPSW-----RDHSVEPLRD-----PNPSDLLENDDSVFSKRAKLEIDEK 515
DB 2547 SLOTEBQWEGAGRGSGIKRHRMPLRDACEPSGDPSTVARIADSSVQTD-----DEE 2599
QY 516 RKRKMDIQRIREQR 529
DB 2600 GEGRYLVTRRRRTR 2613

RESULT 7

T42567

tegument protein 24 - equine herpesvirus 4 (strain NS80567)

C:Species: equine herpesvirus 4
A:Variety: strain NS80567
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42567
R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
J. gen. Virol. 79, 1197-1203, 1998
A:Title: The DNA sequence of equine herpesvirus-4.
A:Reference number: Z22173; MUID:98264497; PMID:9603335
A:Accession: T42567
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3534 <TEL>
A:Cross-references: EMBL:AF030027; NID:92605950; PIDN:AAC59539.1; PID:92605967
A:Experimental source: strain NS80567
C:Genetics:
A:Note: 24
C:Superfamily: varicella-zoster virus gene 22 protein

Query Match 5.98%; Score 192; DB 2; Length 3534;
Best Local Similarity 20.5%; Pred. No. 0.19; Mismatches 158; Conservative 88; Mismatches 271; Indels 252; Gaps 36;

QY 5 SAVFKAAAPAGGPEQRIDYERAAALGPEDEPGAAEAHFLPRHRLKEPGLASSQ-----P 58
DB 2717 TAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSKP-----AAAPAPSKPAAAP 2771
QY 59 LASOGGSPAPSP-PAGCGGKRGILLPAGAPGQOESWGSVPLPCPPATKQAGIGGE 117
DB 2772 APSKPAAPAPSKPAAAPAPSKPAAAPAPSKP-----AAAPAPSKPAAAPAPSK 2820
QY 118 PAAAGA-----GSEPRPYQAVLPQTGSLVAAAEPTPWAGDGGASPA-----ATASDP 169
DB 2821 PAAAPAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSKP-----PAAAPAPSKPAAAPAP 2872
QY 170 AGPPPLPLPGPPPLPATAGTLASSEGKSMKSPGLGGGSGASASQACILKILLLQ 229
DB 2873 SKPAAAPAPSKPAAAPAPSKPAAAPAPSK-----PAAAPAPSKPQNTLVA 2917
QY 230 LDLEQOQOOL--QAKEKEIEELKSERDITLARIEMRRMOLVKNKNEKHKILPGYEETEETELSE 281
DB 2918 IVAKDQAKDKAQKQAKQAKDQAKDQ-----AQDQAKDQAKDQAKQAKQAKDQAKDQD 2972
QY 282 -----LPQYETEE-----RETELSEKIKLECOPELSETQTLIPKPFSC 322
DB 2973 LTKQSNPAKFTGETPLPNTSPSGAVPENTPLDDFIDAVPE-----NTPLPDD----- 3024
QY 323 GSGGKHKRKSPPGSGSTFRKTPVKTLAPESKATKTP-----KHPPIKEPCGSLSETVCKR 379


```

QY 77 KGRGLLIPAGAPGQOEESWGSVPLPCPPPATKAGIGEPAAAGCSEPRKYQAVLP 136
D 66 -----PFGKRP-----FDPNMP-PMPP-----GCIPPMGPPLQRPFFM----- 100
QY 137 IQGSLVAAAKEPTPMAGDKGASPAATASDPAPPLPLPGPPPLPATPTACTLAASE 196
D 101 -----PPPM-----SMPPPMMPGMPVTAATGTPALPT-----E 134
QY 197 GRMKSRKSPGLGGGGSGASS-----QAACLKQILLLDLLEQOQOQLOA 242
D 135 EIWE-NKTPDGKYYNATRESAMTKPDGVKIQGSELPLMLAAQVQAQVQAQVQA 193
QY 243 KEKIEELKSERDPLLARIEMERRMOLVKDNEKRRKLFQGYETEBRETELSEIKL 302
D 194 QAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 249
QY 303 ECOPEL--SETSOPLPKPFGSCGRSGHKHKKSPFGSEKTPVKTLAPERSVKTKTPK 360
D 250 QVQAQVAGASTPTTSSAP-----AVSTTSSTP-----SSTTSTTTTAT-- 290
QY 361 HSPKEPCGSLSETVCKRELRSQETPEK-----PRSSVDTPPRLSTP--OKGPTPK 413
D 291 -----SVAQTVSTPTTQD--TPSSAVSVATPTVSVSTPARATATPVQTVQPHQT 339
QY 414 -----KAF-----SSEIEDLP--LSTTEM 431
D 340 LPPAVPHSVDPQTALPFPVVPVPPRPVPLPGMPLPGVAMQIVSCPVKTVATTKT 399
QY 432 YLCRHQHPSPPLPLRSSPKKEETVARCLMPSSVAGETSVLAVPSMRDHSV----- 483
D 400 GVLPGMAPPIVPM-----IHQVVAIAASPATLAGT--AASETEIKTADGKYYY 448
QY 484 -----BPLRDPNPSDLLENLDVSFSKRAKLELDEKRR 517
D 449 NNRTLESTWEXPKQELKEKEKLEKIKPEKPESEPL-----PMTEEDP 494
QY 518 KRMDIOKIREGRILQRLRMKYGKGIQSEPEVTSFPEPDVESLMTPELV----- 572
D 495 KEPIKIKIEK--PKKEEMTEEKAQAKAPVATAPIG-----TPMCVWMTGDE 542
QY 573 -VAGRPPLKLT 583
D 543 RVEFYNPTRLIS 554

```

RESULT 10

```

A43427
neurofilament triplet H1 protein - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 10-Dec-1999
C:Accession: A43427
R:Soppel, D.R.; Beasley, L.L.; Willard, M.B.
J. Biol. Chem. 267, 17354-17361, 1992
A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polypep
A:Reference number: A43427; MUID:92381055; PMID:1512270
A:Accession: A43427
A:Molecule type: DNA
A:Residues: 1-606 <SOP>
A:Cross-references: GB:M4315; NID:g164990; PIDN:AA57152.1; PID:g601930
A:Experimental source: brain
A>Note: Sequence extracted from NCBI backbone (NCBIN:112010, NCBI:P.112011)
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

```

```

Query Match 5.98; Score 189.5; DB 2; Length 606;
Best Local Similarity 20.48; Pred. No. 0.04; Mismatches 250; Indels 127; Gaps 15;
Matches 114; Conservative 67;

```

```

QY 4 RSAVFAKAAAP-----GPNPQRLDYERAAA-----LGPDEDEGAEAHPLPHRK 51
D 107 KSPYKEAKSPAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEKSPKAEK 163
QY 52 LKEBGPPLASSGGSPAPSPAGCGGKRGRLIPAGAAPGQOEESWGSVPLPCPPATKQ 111

```

```

D 164 AKSPAEEKSPKAKSPVKEAKSPKAKSPVKEAKSPKAKSPKAKSPV-----KE 216
QY 112 AGIGEPAAAGAGCSPRPKYQAVLPIQGSILVAAAKEPTWAGDGAASPAATASDPAG 171
D 217 EAKSPKAKSPVKEAKSPVKEAKSP-----EKAESP-----EKEAKSPAEEKSPEKA 264
QY 172 PPPLPLPGPPLPATATAGTLAASEGRMKSRKSPGLGGGGSGASSQAACLKQILLLDLD 231
D 265 KSPKAKSPVKEAKSPKAEKSPKAKSPVKEAKSP----- 299
QY 232 LIEQOQOQLOAKKEIEELKSERDPLLARIEMERRMOLVKDNEKRRKLFQGYETEER 291
D 300 --EKAKSPVKEAKSPKAKSP-----VKEAKSPKAKSPVKEAKSPKAKSPV 349
QY 292 EETLSERKILCECPLESETSOPLPKPFGSCGRSGHKHKKSPFGSEKTPVKTLAPERSVK 351
D 350 EAKSPKAKSPVKEAKSPKAKSP-----KAKSPV--KEAKSPKAKSPV 397
QY 352 SKVTKTPKHSPIKEPCGSLSETVCKRELRSQETPEKPRSSVDTPPRLSTPQKGPSTHP 411
D 398 EAKSPKAKSPVKEAKSPKAEKSPKKEEVKPKVEBETAPAPKVEKDSKKDEAPK 457
QY 412 KE-----KAFSEIEDLPYLSTTEYLCRHQ----- 439
D 458 KEAPKPAVERKRESTAAKDEADKRAAPAKMEGKEAKPKETEVAKKEPEDAKAKE 517
QY 440 PSPPLPRLSSPKKEETVARCLMPSSVAGETSVLAVPSMRDHSVPLRDPNPSDLLENLD 499
D 518 PSKTEKEPEKPKKEETPA-----APYKKA-----KEAKKPEKPK--TEAKKED 563
QY 500 DSVFSKRAKLELDEKRR 517
D 564 DKALSKPEKSKPKTEKAEK 581

```

RESULT 11

```

T28770
hypothetical protein W03D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T28770
R:Kohlring, T.; Wohlmann, P.
submitted to the EMBL data library, June 1997
A:Description: The sequence of C. elegans cosmid W03D2.
A:Reference number: Z20519
A:Accession: T28770
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-539 <ROH>
A:Cross-references: EMBL:AF000298; PIDN:AAC48255.1; GSPDB:GN00022; CESP:W03D2.1
A:Experimental source: strain Bristol N2; clone W03D2
C:Genetics:
A:Gene: CESP:W03D2.1
A:Map position: 4
A:Insertions: 40/3; 88/3; 115/3; 146/3; 173/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homolog

```

```

Query Match 5.88; Score 189; DB 2; Length 539;
Best Local Similarity 24.88; Pred. No. 0.037;
Matches 109; Conservative 23; Mismatches 132; Indels 176; Gaps 21;

```

```

QY 9 KAAAPAGGNPQRLDYERAAALGPEDEGAEAHPLPHRKLEKGPPLASSGGSPA 68
D 249 KRQAPRAGSP-----PPPP-----KSPPLAGS--GSP 277
QY 69 PSPAGCGGKRGRLIPAGAAPGQOEESWGSVP-----LPCPPATKAGIGEPAAAGA 123
D 278 PPPA-----AGSPPPRT-----GSPPPPTGSPPPPA-----GSPPPRA 315
QY 124 GCSPRKRYQAVLPIQGSILVAAAKEPTPMAGDKGASPAATASDPAGPPP----- 174
D 316 GSPPPPPPPRGSP-PTGSL-----PPQA-----GSPPPAGTGSPP--PPPRQKROAPER 363

```

```

QY      175  -----LPGLCPPLAATAGTATLASEGKMKMKRSPILGG--GGGSGASSQAQACIKILL 227
Db      364  SPTTSGPPPGSGPPTGTPPPGCGPKSKSESSSE--KGGPPGCGRGGCGRP----- 411
QY      228  LQDLLEQ0000LQAKKEIEELKSERDTLLARIEMERMOLVUKDNEKERHKLFGQYE 287
Db      412  -----RKSSSE-----S 418
QY      286  TEERETELSEKIKLECOPELSETSQLPAPKFGSCGSGKHKRRKSPGSTERTPYKLL 347
Db      419  SERSEPR-----GPRSPPGSGPPTGSPPTGPRPG--SPTGSPPLGLP---- 462
QY      348  APESFVKTKTPKHSPIKEEPCGSLSEYCKR---ELRSQETPEKPRSSVDTPRLSTP 403
Db      463  ----SHQKQADREDRTGSPPTGSPPTGPRPHNGCGPKGKSESESRGPRRRSPPTGSP 518
QY      404  QKGSP--HPKKAFASSIED 422
Db      519  TGSPTGAPPKKGFASLPD 538

```

RESULT 12

eyelid - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: J13049
 R:Treisman, J.E., Luk, A., Rubin, G.M.: Heberlein, U.
 submitted to the EMBL Data Library, March 1998
 A:Reference number: Z17592
 A:Accession: J13049
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2715 <TR>
 A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
 C:Genetics:
 A:Gene: eld
 A:Cross-references: FlyBase:FBgn0003013
 C:Function:
 A:Description: could act as a transcription factor antagonistic to the wg pathway
 :Keywords: DNA binding

Query Match	5.8%	Score 188;	DB 2;	Length 2715;
Best Local Similarity	21.4%;	Pred. No. 0.22;		
Matches 125;	Conservative 41;	Mismatches 175;	Indels 242;	Gaps 26

QY	11	AAABAGGPEERJLYERAAALGGREDEGAEEANLPRIHRIKKE	-GGPRLASSQ--	GGSP	67	
Db	70	APPPGSGP-----	GGPGRDPAAMVNIHNIHQOQOQDPRRNMQOQDNHSGR		117	
QY	68	APSPAGSGSGKRGILLPRAGA--	APGOEESMGSVRLPCRRPA-----	-TK	110	
Db	118	APPP-----	PGARENHAAGVLEEY--	THLPRIHPRAYGRYHADENMDPRYRG	163	
QY	111	QAGTGE-----	PAAGAGGCSG--	RPKYOAVLPDQ-----	TSLVAAKE	148
Db	164	QRLRGKRRQOQOQYHNRQOQRPQOQRRGGSSPNRRPOQRXIRGGRPGGPTTLMSLLQSSNP			223	
QY	149	PTP---WAGD-----	KGAASPRATASDPRGPRLLRGPRPLAPATTAAGTLAASEGR		198	
Db	224	PRPRONRYANTYDPRQOAAASAAAAAQAQOQGGRRPRGNHGRPR--	ROHQSPRGGOQGG		281	
QY	199	W-----	KSMRKSPD-----	GGGGSGGASSQAALQILLD	230	
Db	282	WAPRRPRYSPOLGSPSOQYRTPRPTLTSKQOQYPRAHGONSQSYSSP-----			329	
QY	231	DLIEDQOQOOLAKEREIEELKSESDTLARIERMERRMOLYUKDKNEKRNKLPQGYETEE			290	
Db	330	--:QOQOQOQOQOQOQA-----	QOQPRGPRPGGPRPT		359	
QY	291	KEEPELSEKILCELOPELSETSQTLP-----	PKPPSCGSGGCHK--	RKS	333	

```

Db      360  GGGPR-----QQNPRTSQSPYRQRYPTTGGRLPAGGSNNRTAVSTHQYDEPNRM 410
QY      334  PFSTSEKTRV\KKLAREFSKVTKPKH-SPIKEPCCSLSETVCKRELNSOETPEKPRS 352
        411  PGSSSPSPGSGHPLRP-----ASPHNVPLOOQP-----PPPHV 445
Db      393  SVDTPLRLSPQGPSTHPEKAFSELEDLRYLSTEMYLCSMHNORRPSPLRESPEK 452
        446  SAGGPR-----PSSSP-----GHASPSPQSPQASPS 473
QY      453  KEETVARCLMPSSVAGETSVLAVPSWRDHSVEPLRPNPSDL 495
        474  HQELIGNSNDSSGSHGSM-----GSPFPTGPNQOVM 508

```

RESULT 13

hypothetical protein C18H7.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: J33110
C:Rita-Wollam, A.; Fronick, W.
A:Submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid C18H7.
A:Reference number: Z21284
A:Accession: J33110
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-460 <TI>
A:Cross-references: EMBL:AF067607; PIDN:AAC17641.1; GSPDB:GND0022; CESP:C18H7.3
A:Experimental source: strain Bristol N2; clone C18H7
C:Genetics:
A:Gene: CESP:C18H7.3
A:Map position: 4
A:Introns: 84/1
C:Superfamily: Phasocelus glycine-rich cell wall protein 1.8

Query Match	5.88;	Score 186.5;	DB 2;	Length 460;
-------------	-------	--------------	-------	-------------

Matches 81; Conservative 14; Mismatches 111; Indels 73; Gaps 12;

QY	12	AAAGAGPPERLDLYEFAALG-----GPEDEFGAA-----EAHFLPRHRLKEPG	56
Db	152	AGPAGPAGNMGADGE-AGAEPPAGEAGPPGPPGPKRGAAAGSDGEAG-AGCTTNAAG	209
QY	57	PPLAASOGSGPAPSPAGCGCGKGRGLLPLAG-----AAPQOEESMGSVP----	101
Db	210	EPSPAGPAGPPGPGNGSAGAGSGPGEPPGAPGPPGPKGDEAGAGAKGDDAGADAPGTD	269
QY	102	--LPCEP-PAITKQAGTAGEPAAA-----GAGCSPPREKYQAVLPQ	138
Db	270	AYCCPPRSAILGAGGAEPPAGAPAAPEAAAAPEAAPEAAPEAAPEAGAGAEPPAGAAADAA	329
QY	139	TGSLVAAAKEPTFMAGDKGGAASPATASDPAGPPLP-----GPPAPPT	186
Db	330	AAAPAEAPAEAPAAAGAGGAEPPAGAPADAAAAPEAAPAEAPAAEGAGGAEPPAGAA	389
QY	187	ATAGTAAASGGRKSKRSPL--GGGG--SGASSQA	220
Db	390	APDAAAAPEEA-AAPEAAPAAEGAGGAEPPAGAPAEA	427

RESULT 14

hypothetical protein B7F21.40 [imported] - *Neurospora crassa*
C.Species: *Neurospora crassa*
C.Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C.Accession: T51023
R.Schulze, U., Aign, V., Hohnesiel, J., Brandt, P., Fartmann, B., Holland, R., Nyakatu
submitted to the Protein Sequence database, July 2000
A.Reference number: Z5286
A.Accession: T51023
A.Status: preliminary

A: Molecule type: DNA
 A: Residues: 1-2649 <SCH>
 A: Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
 A: Experimental source: BAC clone B7F21; strain OR7A4
 C: Genetics:
 A: Gene: NCSP:B7F21.40
 A: Map position: 6
 A: Introns: 1619/3; 2584/1

Query Match 5.7%; Score 184.5; DB 2; Length 2649;
 Best Local Similarity 23.2%; Pred. No. 0.32;
 Matches 118; Conservative 60; Mismatches 171; Indels 159; Gaps 28;

```

QY      34 PEDPFGAAEAHFLPR---HRKL-----KEGEPPLASQGGSPAPSGCGGKGLLPA 85
DB      1656 PTPSAGPRKRYDVSSSGHPRPLAAAEAEPEOP--AKSEAAAPRANQPF-----SHFYQTPI 1708

QY      86 GAAPGQOEESMGSVPLDPC-PPATKQAGIGGEPAAAGAGCGSPRKKQAVLPITGSLVA 144
DB      1709 QASPVLTIVORSFTVPAPLPRA-----PVAASASAPSPQ-----TPSRPGAVS 1755

QY      145 AAKPTPMAGDKGGAASPAAT--ASDPAGPPLPLPGPP-----PLAPATAG 190
DB      1756 QTMSPVPHP-----LRQPTATFVTEREG--EPVVSQPPAQTQHPYRISQKTPAPVSSS 1809

QY      191 TLAASEGKMSKRSPLGGGGSGSQAACLKQILLDLIEQOQOOLQAKKEITEEL 250
DB      1810 MPASSEA-----MPRSAGMIPRANRPTLLSQ-----QHLEIREVEI 1846

QY      251 KSEEDTLARLEREMRMQVKKKNEKRRHKLFGQYETEEREETEELSEKIKLEQPELSE 310
DB      1847 REGGEROI--RHEPLQPREQPPMERPRMRK-----QEPE-----QPLHHD 1886

QY      311 TSQT--LEPKPFSCGSGKGRKSPFESTERTKTPVKKLAEFSKVTKTPKHSPIKEEPC 369
DB      1887 PFOFMFMPQ-----RSMQPRAEAPMG---RQDPRRMAP--SAPQSYTP--PIQAPV 1933

QY      370 GS--LSEYCKRELRSQETPEKPRSSVDTPPRLST-----PQ 404
DB      1934 RNLLSESV-----PPQRTPOLSPAMERPVPSPQRPMPSTMOEYSAQVPPVPPQ 1988

QY      405 KGPSTHPEKKAFFSEIDLYLSTEMYLCRMHQPSPPLPLRSSPKKEETVARCLMPS 464
DB      1989 APAPIPEPKKVSISFILN-----DDNPAPAP---AKRVNDVAS--MPR 2030

QY      465 SVAGETSVLAVPSWRHDSVEPLRDPNPS 492
DB      2031 AASTST-----PPQQMSARPPQPPPT 2053
  
```

RESULT 15

T00273
 hypothetical protein KIAA0595 - human (fragment)
 C: Species: Homo sapiens (man)
 C: Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C: Accession: T00273
 R: Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
 DNA Res. 5, 31-39, 1998
 A: Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
 A: Reference number: Z14086; MUID:98290545; PMID:9628581
 A: Accession: T00273
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-1520 <NAG>
 A: Cross-references: EMBL:AB011167; NID:93043713; PIDN:BAA25521.1; PID:93043714
 C: Experimental source: brain
 C: Genetics:
 A: Note: KIAA0595

Query Match 5.7%; Score 183.5; DB 2; Length 1520;
 Best Local Similarity 20.6%; Pred. No. 0.2;
 Matches 153; Conservative 64; Mismatches 216; Indels 311; Gaps 34;

```

QY      13 APAGNPEORLDYERAAALGGPEDEPGAAEAHFLPRHRRKLEKGPPLASSQGGSPAPS-- 70
DB      681 SPASSEPEPVSKPVASS---PTEQVPSQEMPLAR-----PSPVQSVPAVPTPPSM 731

QY      71 ----PACGGKGRGLLPPAGAAGQOEESMGSVPLDCCPPATQAGIGEP-----AAA 121
DB      732 SAALPPAGGLGMPSSLP---PPLOPPSLPLSGVPLPDFTHYALPLPSMPCPHVSPS 788

QY      122 GAGCSPRKYQAVLPITGSLVAIAAKEPT---PWAGDKGAASPAATADPAGPPLPLP 178
DB      789 GYCELP---PPVPLVSGTPGAVAPPTCSVPA-----PPAPV- 826

QY      179 GPPPLATATAGTLAASEGKMSKRSPLGGGGSGSQAACLKQILLDLIEQOQO 238
DB      827 --SPYSSTCTYG-----PLGMPG----- 843

QY      239 QLAKEKEIEELKSERDTLLARLEREMRMQVKKKNEKRRHKLFGQYETEEREETEELSE 298
DB      844 ----- 843

QY      299 KIKLECPLELSESTPLPKPF---SCGRS---GKGHRKSPFESTERTKTPVKKLAP--- 349
DB      844 ----PQNAPEWSTVPPPLPPASIGRAVPQPKMSRGTGAPRPENVLPLSAPPLSL 896

QY      350 -----EFSKVKTKT-----PKH-----SP--IKEPCCS-----LSEYCKR- 379
DB      897 GLPCHGAPQTEPTVEVKVPVAPSPHPRKHKSALVQSPOMKALACVSAEGTVVEPASERL 956

QY      380 ----ELRSQETPEKPRSSVDTPPRLSTPKGPSTHKE-----KA 415
DB      957 KPETOETRPKEKPLPATKAVPTPQSTVPLPAVHARLRKLSFLTPTQSGSEDVQA 1016

QY      416 FSSSEI---BDL--PYLSTEMYLCRMHQPSP---LPLRESS---PKKEETVARCLMP 463
DB      1017 FISELTGEASDLSLLEQFEKSEAKKCEPPAPADSLAVNGSGVDIPQEKRLDRQAP 1076

QY      464 --SSVAGETSVLAVPS--WRDHSVEPL---RDP-----NPSDLLENLDDSVFSKR 506
DB      1077 ELANVAGLTTPPAPPHQMLKPLAAVSLAKAKSPKSTAQGTGTLKPEGVTE-----AKH 1129

QY      507 HAKLELDE-----KRRKMDIORIREQRIQ 532
DB      1130 PAAVRLQEGVHGSRVAVGSGDHDYCYRSTPPKKMPALVIPYEGSMNVKRDQITIKP 1189

QY      533 RLQRLMYKKKGIOSEPEVTSFPPEPD--VESLMTPLPLVYAFGRPLKLPDQNEFLP 590
DB      1190 VLSL-----GPAAPPPCIAASREPLDHTSSQADPSAPCLA---PSSLSP----- 1234

QY      591 WLDERSKRLLEIOKKQTPHRTCKR 614
DB      1235 ---EASPCRDMDMTRTPEPSAKQ 1255
  
```

Search completed: July 14, 2003, 11:51:29
 Job time : 48 secs

XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 11: Page 1153-1154; 1425pp: English.
 XX
 CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytostatic, cardioactive, and
 CC immunomodulatory; muscular active general; vulnerable; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 XX Sequence 281 AA;
 SQ
 Query Match 44.7%; Score 1446; DB 21; Length 281;
 Best Local Similarity 100.0%; Pred. No. 3.8e-89;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 181 PPLAPATAGTTLASBGRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQOQQL 240
 DB 1 PPLAPATAGTTLASBGRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQOQQL 60
 QY 241 QAQEKIEELKSRDTLLARIEMERRMOLVKNDKEKRLKFGYETERETELSEKI 300
 DB 61 QAQEKIEELKSRDTLLARIEMERRMOLVKNDKEKRLKFGYETERETELSEKI 120
 QY 301 KLECOPELSTOTLPKPPSCGSGKGRKSPFSTERTKTPVKKLAPFSVKTKTPK 360
 DB 121 KLECOPELSTOTLPKPPSCGSGKGRKSPFSTERTKTPVKKLAPFSVKTKTPK 180
 QY 361 HSPKKEPGCSLSETVCKRELNSQETPEKRRSSVDPPLRSTPOKGPSTHPKKAASSEI 420
 DB 181 HSPKKEPGCSLSETVCKRELNSQETPEKRRSSVDPPLRSTPOKGPSTHPKKAASSEI 240
 QY 421 EDLPYSTETMYLCRWHOPPPSPPLPRESSPKKEETVA 458
 DB 241 EDLPYSTETMYLCRWHOPPPSPPLPRESSPKKEETVA 278
 DB
 RESULT 2
 AAB42665
 ID AAB42665 standard: Protein; 269 AA.
 XX
 AC AAB42665;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2429, polypeptide sequence SEQ ID NO:4858.
 XX
 XX Human ORFX ORF2429, polypeptide sequence SEQ ID NO:4858.
 KM Human; Open reading frame: ORFX; detection: cytostatic; hepatotropic;
 KM vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antineoplastic;
 KM antiviral; antibacterial; antifungal; antineumatic; antihypertoid;
 KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;

KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR N-P-SDB; AAC76874.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 11: Page 4046; 5507pp: English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertoid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 269 AA;
 Query Match 39.4%; Score 1277; DB 21; Length 269;
 Best Local Similarity 94.3%; Pred. No. 7.7e-78;
 Matches 247; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 197 GRKMSKRSPLGGGGSGASSQAACLKQIILLQDLIEQQOQQLQAKKEIELEKSEKRT 256
 DB 5 GQMEYEYEEFSSRQWRLGSSSSQAACLKQIILLQDLIEQQOQQLQAKKEIELEKSEKRT 64
 QY 257 LLARIEMERRMOLVKNDKEKRLKFGYETERETELSEKIKLECOPELSTOTLP 316
 DB 65 LLARIEMERRMOLVKNDKEKRLKFGYETERETELSEKIKLECOPELSTOTLP 124
 QY 317 PKPFGSGRGSGKGRKSPGSTERKTPVKKLAPFSVKTKPKHSPKKEPGSLSETV 376
 DB 125 PKPFGSGRGSGKGRKSPGSTERKTPVKKLAPFSVKTKPKHSPKKEPGSLSETV 184
 QY 377 CKRELNSQETPEKRRSSVDPPLRSTPOKGPSTHPKKAASSEIEDLPYSTETMYLCRW 436
 DB 377 CKRELNSQETPEKRRSSVDPPLRSTPOKGPSTHPKKAASSEIEDLPYSTETMYLCRW 436

Db 185 CKRELRSQETPEKERRSSVDTPTPRLSTPQKSGSTHPEKKEAFSSIEDLPYLSTEMYLCRW 244
 QY 437 HOPPSPLPLRESSPKKEETVA 458
 Db 245 HOPPSPLPLRESSPKKEETVA 266

RESULT 3
 AAU30711
 ID AAU30711 standard; Protein: 195 AA.
 AC AAU30711;
 DT 18-DEC-2001 (first entry)
 DE Novel human secreted protein #1202.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX

OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HXSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20: Page 334; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression, and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33504 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 195 AA;

Query Match 11.7%; Score 377.5; DB 22; Length 195;
 Best Local Similarity 46.3%; Pred. No. 1.1e-17;
 Matches 94; Conservative 14; Mismatches 72; Indels 23; Gaps 8;

QY 10 AAAAPAGGNDEQRADYERAAALGPEDEPGAAEAHFLPRHRKLK-EPGPIASSQSGSPA 68
 Db 1 AAAAPAGGNDEQRADYERAAALGPDGRAMGSRPLPPAPXAGAGPGRPPRRAGSPA 60
 QY 69 PSPAGC-GGKGRGLLPAGAAP--GQOESWGSGSVLPCCP--PPATKAGIGG-----E 117
 ||||| ||||| ||: | :| ||| | : |||

Db 61 PSPAGCGGKGGGLVTPGRGGRAPRAAGREL----AVRCPCPVRRPPSPKALGSLDPOE 116
 QY 118 PAAAGAGCSPPRKQAVLPDTGSLVAAAEPTWADKGAAPATASPPAPPPPL 177
 |||| | : ||||| | : | : ||||| |
 Db 117 PAA-----PGPSIRPVLPDTGSPWRPRSLRVLGTGVRPPPLPPPPDPAGPPPL 171
 QY 178 PGPPPLAPATAGTL--AASEGR 198
 ||| | | : |||
 Db 172 GPHPSRPPPTGPPWRAPRADGR 194

RESULT 4
 AAE20839
 ID AAE20839 standard; Protein: 720 AA.
 AC AAE20839;
 DT 01-JUL-2002 (first entry)
 DE Human gene 7 encoded secreted protein fragment, SEQ ID NO:101.
 XX
 KW Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
 KW immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
 KW rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
 KW sezar syndrome; Gaucher's disease; neurological disease; cardiac arrest;
 KW Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
 KW cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
 KW thrombosis; ocular disorder; corneal infection; wound healing; cardiac;
 KW vascular; thrombolytic; cytosstatic; nootropic.
 XX
 OS Homo sapiens.
 XX
 PN WO200218435-A1.
 XX
 PD 07-MAR-2002.
 XX
 PF 17-JAN-2001; 2001WO-US01567.
 XX
 PR 28-AUG-2000; 2000US-228084P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 PI Ni J;
 XX
 DR WPI: 2002-281060/32.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
 PT cardio-/cerebrovascular disorders and multiple sclerosis -
 XX
 PS Disclosure: Page 18-19; 504pp; English.
 XX
 CC AD33237-AD33280 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
 CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
 CC infections, anaemia, rheumatoid arthritis and multiple sclerosis),
 CC cancers and hyperproliferative disorders (e.g. melanoma, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
 CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
 CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
 CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
 CC infections). Secreted proteins of the invention can also be used to

CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 720 AA:

Query Match 7.0%; Score 226.5; DB 23; Length 720;
 Best Local Similarity 23.1%; Pred. No. 7.5e-07;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYERAALGDEPERGAANEHLRHRKLEKRPPLASSGGS-----PAPSPAG 73
 DB 50 PKREELIQ-AQIKGPVDPVGRDGEKAEPEADLRPGGIAVPVGEARHREPPVPHDKV 108
 QY 74 CGSGKRGILLP-----AGAPG-----QOESWGSGSVPLPCPPA 108
 DB 109 VVDEGQDREVEENKPPSRHAGKAPGVGOMAPLPDSEREKQEPGEGVGKRPQQA 168
 QY 109 TKQAG-----IGEPAAAGA-----GCSPPKXYQAVLPLOTGSLVAAKE 148
 DB 169 LEEAGGLPDPKQVVPADQGPVQPAKEDLGPDRGRLHRRP--QAVLSQONGLANVAGGE 226
 QY 149 PT---PWAGDKGGAASPAATASDPAPPLPL--PGPPLPATATAGTLAASGKRWKSMR 203
 DB 227 KAGGPPPGNAADDTQOPADSDHGGKPLPAKRPAGPLP-----EPRE 273
 QY 204 KSPILGGGSGGSSGQ-----AALCKQILLDLLEQQOQQLAKEKEIEELSERDTL 257
 DB 274 QRVVERAGGQAQSLAEAGRAMLDHVALVLY-IEQOQVQ-----KRLDQEK 324
 QY 258 LAIEKERRMQLVKKDKNEKERHKLPGVETEBRETELSEKIKLEC-QDELSETQTLR 316
 DB 325 LAVIEQDKHKEIHOOORDEDEDKPR---QVEVHOEPGAVPVPGQEAPEKARETEVNLPR 379
 QY 317 PKPFSGSGSGKHKRSPGSTERKTPVKKLAFESVVKTKPKHS--PIKEBSCSLSE 374
 DB 380 PLPLD-----FVLAPG--RPAPSDL-----NORSLEHSGPVRGAPAG--- 418
 QY 375 TWCKRELRSQETPEKPRSSVDPRLSTPGKPSHPKE---KAFSSEIEDLPLYLSTTEM 431
 DB 419 -----PPDGGPDTERRAAGKLRGQKNAARFAGTVK 451
 QY 432 YLCRMHQPPSPPLARESSPKKE 454
 DB 452 ELPKGPVOYVPPDAREAGPEE 474

RESULT 5

ID AAE20838 standard; Protein; 766 AA.

AC AAE20838;

DT 01-JUL-2002 (first entry)

XX Human gene 7 encoded secreted protein fragment, SEQ ID NO:100.

KM Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
 KM immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
 KM rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasia;
 KM Sezary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
 KM Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
 KM cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
 KM thrombosis; ocular disorder; corneal infection; wound healing; cardiant;
 KM vascular; thrombolytic; cytostatic; nootropic.

OS Homo sapiens.
 XX
 XX WO200218435-A1.
 XX
 XX 07-MAR-2002.
 PD
 XX
 XX 17-JAN-2001; 2001MO-US01567.
 XX
 XX 28-AUG-2000; 2000US-228084P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fischella M;
 PI Ni J;
 XX
 XX WPI; 2002-281060/32.

PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
 PT cardio-/cerebrovascular disorders and multiple sclerosis -
 PS Disclosure; Page 18; 504pp; English.

CC AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
 CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
 CC infections, anaemia, rheumatoid arthritis and multiple sclerosis),
 CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
 CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
 CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
 CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
 CC infections). Secreted proteins of the invention can also be used to
 CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 766 AA:

Query Match 7.0%; Score 226.5; DB 23; Length 766;
 Best Local Similarity 23.1%; Pred. No. 8.1e-07;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYERAALGDEPERGAANEHLRHRKLEKRPPLASSGGS-----PAPSPAG 73
 DB 96 PKREELIQ-AQIKGPVDPVGRDGEKAEPEADLRPGGIAVPVGEARHREPPVPHDKV 154
 QY 74 CGSGKRGILLP-----AGAPG-----QOESWGSGSVPLPCPPA 108
 DB 155 VVDEGQDREVEENKPPSRHAGKAPGVGOMAPLPDSEREKQEPGEGVGKRPQQA 214
 QY 109 TKQAG-----IGEPAAAGA-----GCSPPKXYQAVLPLOTGSLVAAKE 148
 DB 215 LEEAGGLPDPKQVVPADQGPVQPAKEDLGPDRGRLHRRP--QAVLSQONGLANVAGGE 272
 QY 149 PT---PWAGDKGGAASPAATASDPAPPLPL--PGPPLPATATAGTLAASGKRWKSMR 203


```

Db      273 KAKGPPPPNAGDTGQPAEDSDHGKPLPAEKPAFGPLP-----EPRE 319
Qy      204 KSPFGGGGSGASSQ-----AACLKQILLDLILQOOQOLQAKKEIELKSERDTL 257
Db      320 QRDVERAGGNQASQLEBAGRAEMLDHVALLVQV-IRKQVQVQ-----KRLDQOEKL 370
Qy      258 LARIERERRMQLVKKNKEKRNKLFQGYETEREETELSEKIKLEC-QPELSETSOTLP 316
Db      371 LAVIEBQHKELHQQRDEDEKPR-----QVEVHQDEGAANVPRQDEAPEGKARETVENLP 425
Qy      317 PKPSCGSRSGKHKRSPFGSTERKTPVKLAPESVKTKTPKHS--PIKEPCGSLSE 374
Db      426 PLPLD-----PVLRAAPG--RPAPSQDL-----NQRLSHSGPVRGDPAG---- 464
Qy      375 TVCKRELRSQETPEKPRSSVDTPPRLSTPOKGPSTHPKE--KAFSSEIEDLPYLSTTEM 431
Db      465 -----PPDGGPTTEPRAAGKLRDQKDAAPRAAGTVK 497
Qy      432 YLCRMHQPSPPLPRLRESSPKKE 454
Db      498 ELKPGPEQVPVDPDAREAGPEE 520

RESULT 6
ABG64675
ID ABG64675 standard; Protein: 973 AA.
XX
XX ABG64675:
XX
XX 27-AUG-2002 (first entry)
XX
XX Human albumin fusion protein #1350.
XX
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
XX human serum albumin; HSA; cancer; reproductive disorder;
XX digestive disorder; immune disorder; endocrine disorder;
XX haematopoietic disorder; neural disorder; connective disorder;
XX cytostatic; antileukemia; antineoplastic; antitumor;
XX immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
XX neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
XX osteopathic; antiarthritic.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX WO20017137-A1.
XX
XX 18-OCT-2001.
XX
XX 12-APR-2001; 2001WO-US11988.
XX
XX 12-APR-2000; 2000US-229358P.
XX 25-APR-2000; 2000US-199384P.
XX 21-DEC-2000; 2000US-256931P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Haseltine WA;
XX
XX WPI; 2002-010886/01.
XX
XX New fusion protein for treating disease e.g. diabetes comprises an
XX albumin fused to a therapeutic protein -
XX
XX Claim 1; Page 1418-1421; 2102pp; English.
XX
XX The present invention relates to albumin fusion proteins comprising a
XX therapeutic protein X and human albumin (HA, also known as human serum
XX albumin, HSA). The proteins are useful for treating a disease or
XX disorder that may be modulated by therapeutic protein X. The albumin
XX extends the shelf-life of protein X, and may increase its biological
XX in vitro/in vivo activity. The protein is useful for treating and

```

```

CC diagnosing disorders such as cancer, reproductive disorders, digestive
CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC (e.g. diabetes), haematopoietic disorders, neural disorders
CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
CC fusion proteins of the invention.
XX
XX Sequence 973 AA:
XX
XX Query Match 7.0%; Score 226.5; DB 23; Length 973;
XX Best Local Similarity 23.1%; Pred. No. 1.1e-06;
XX Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

Qy      19 PEORLYERAALGPEDEPGAAHFLPRHRIKKEGPPPLASSGGS-----PAPSPAG 73
Db      303 PKERELEQ--AQIKGPDVVGREDGKAPAEBAQDQDRGQGIAPVGEAHNHEPPVPRDKV 361
Qy      74 CGGKGRGLLP-----AGAAPG-----QOESMGSGVPLPCPPA 108
Db      362 VDEGQDRENVPEENKPPSRNAGKAPGVQGMAPPLDSEKKEQPEQEGVKRPGQAQA 421
Qy      109 TKQAG-----IGEPAAAGA-----GSPREKYOAVLPYQTSILVAAKE 148
Db      422 LEAGDLPEDPQKVPKPADGQPAVQPAKEDLGPDGRGLHPR--QAVLSEQNGLAVGGE 479
Qy      149 PT---PWAGDKGAASPAATASDPAGPPPLP--PGPPPLAPYTAGTLAASEGRKMSMR 203
Db      480 KAKGPPPPNAGDTGQPAEDSDHGKPLPAEKPAFGPLP-----EPRE 526
Qy      204 KSPFGGGGSGASSQ-----AACLKQILLDLILQOOQOLQAKKEIELKSERDTL 257
Db      527 QRDVERAGGNQASQLEBAGRAEMLDHVALLVQV-IRKQVQVQ-----KRLDQOEKL 577
Qy      258 LARIERERRMQLVKKNKEKRNKLFQGYETEREETELSEKIKLEC-QPELSETSOTLP 316
Db      578 LAVIEBQHKELHQQRDEDEKPR-----QVEVHQDEGAANVPRQDEAPEGKARETVENLP 632
Qy      317 PKPSCGSRSGKHKRSPFGSTERKTPVKLAPESVKTKTPKHS--PIKEPCGSLSE 374
Db      633 PLPLD-----PVLRAAPG--RPAPSQDL-----NQRLSHSGPVRGDPAG---- 671
Qy      375 TVCKRELRSQETPEKPRSSVDTPPRLSTPOKGPSTHPKE--KAFSSEIEDLPYLSTTEM 431
Db      672 -----PPDGGPTTEPRAAGKLRDQKDAAPRAAGTVK 704
Qy      432 YLCRMHQPSPPLPRLRESSPKKE 454
Db      705 ELKPGPEQVPVDPDAREAGPEE 727

RESULT 7
AAE20799
ID AAE20799 standard; Protein: 973 AA.
XX
XX AAE20799:
XX
XX 01-JUL-2002 (first entry)
XX
XX Human gene 7 encoded secreted protein HMVBP38, SEQ ID NO:61.
XX
XX Human; secreted protein; gene therapy; human immunodeficiency virus; HIV;
XX immune disease; autoimmune disease; anaemia; multiple sclerosis; cancer;
XX rheumatoid arthritis; hyperproliferative disorder; melanoma; neoplasm;
XX seazary syndrome; Gaucher's disease; neurological disease; cardiac arrest;
XX Alzheimer's disease; Parkinson's disease; Charcot-Marie-Tooth disease;
XX cardiovascular disorder; cerebrovascular disorder; tachycardia; angina;
XX thrombosis; ocular disorder; corneal infection; wound healing; cardiant;
XX vascular; thrombolytic; cytosolic; neurotropic.
XX
XX Homo sapiens.
XX

```

FH Key Location/Qualifiers
 FT Peptide 1..34
 FT /label= signal_peptide
 FT Protein 35..973
 FT /label= Mature_secreted_protein
 XX WO200218435-A1.
 XX 07-MAR-2002.
 XX 17-JAN-2001; 2001MO-US01567.
 XX 28-AUG-2000; 2000US-228084P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 PI N1 J;
 XX WPI; 2002-281060/32.
 DR N-PSDB; AAD33243.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating e.g. Alzheimer's disease,
 PT cardio-/cerebrovascular disorders and multiple sclerosis -
 XX
 XX Claim 11; Page 451-454; 504pp; English.

CC AAD33237-AAD33280 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE20793-AAE20836 represent the proteins they encode.
 CC AAE20837-AAE20847 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases (e.g. HIV (human immunodeficiency virus)
 CC infections, anemia, rheumatoid arthritis and multiple sclerosis).
 CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-
 CC Marie-Tooth disease), cardiovascular or cerebrovascular disorders (e.g.
 CC cardiac arrest, tachycardia, angina and thrombosis), infections caused
 CC by bacteria, viruses and fungi and ocular disorders (e.g. corneal
 CC infections). Secreted proteins of the invention can also be used to
 CC promote wound healing, maintain organs before transplantation, support
 CC cell culture of primary tissues, modulate differentiation of embryonic
 CC stem cells, induce mesodermal tissue to differentiate in embryos,
 CC modulate mammalian characteristics (e.g. height and weight), modulate
 CC the catabolism, anabolism, energy storage, mental state, biorhythms,
 CC cardiac rhythms, reproductive potential, hormonal levels appetite,
 CC memory and stress. They can also be used as an additive to increase
 CC or decrease storage capabilities and nutritional content of food. The
 CC present sequence represents a human secreted protein of the invention.

XX Sequence 973 AA;

Query Match 7.0%; Score 226.5; DB 23; Length 973;
 Best Local Similarity 23.1%; Pred. No. 1.1e-06;
 Matches 116; Conservative 61; Mismatches 181; Indels 145; Gaps 21;

QY 19 PEORLYTERAALGPEDEPGAALAHFLPRHRLKKEGPPLASSOGS-----PASPAG 73
 DB 303 PKRELEIQ-AQIKGEYVDVGRDCKEAPBEAQLDRGGCIANPVGEAHNHEPPVPHDKV 361
 QY 74 CGKGKGLLP-----AGAPG-----QQESWGSVLPPLCPPPA 108
 DB 362 VVDEGQDREVPENKPPSRHAGKAPGVGOMAPLPDSEEREKQPEQGVGKRPQQAQ 421
 QY 109 TKKAG-----IGGEPAAGA-----GCSPPKIQAVLPLOTGSLVAAKE 148

DB 422 LEEAGDLPEDDPOKVPPEADGQPAVQPAKEDLGPDRGJHPRP--QAVLSEQONGLAVGGE 479
 QY 149 PT---PWAGKKGGAAPAAASPAPRPLPL--PGPPLAPATATGTLAASGRKKSAR 203
 DB 480 KAGGPPPGNAADTQGPAPEDSDHGKPPPAKPPAPGGLP-----EPRE 526
 QY 204 KSPLDGGGGSGGASQ-----AACLKQILLDLLEQQOQOQAAKEKEIEELKSPDTL 257
 DB 527 QRDVERAGGNGAASQLEENGRALMDHAVLLQY-IEQQVQO-----KRLDDQEL 577
 QY 258 LARIERNRRMQLVKDNEKERKHLFGYTEEREEETELEKIKLDC-QPELSETQTLP 316
 DB 578 LAVIEQHKKEIHQORDEDEDKPR-----QVEVHOEPGAAPVPGQAPAEKAREVENLP 632
 QY 317 PKPFCGSRGSKGKRRSPFSTRTKTPVKKLAPEFSKVTKTKHS--PIKEPCCSLSE 374
 DB 633 PLPLD-----PVLAPGQ--RPASQDL-----NORSLHEGPGVGRDPA-- 671
 QY 375 TVCKRELRSQETPEKPRSSVDTPTPLSTPQKGSJHPKE--KAFSEIEDLPYLSTTEM 431
 DB 672 -----PPDGGPTPEPRAAGKLRDQGDADAPRAAGTVK 704

QY 432 YLCRWHPSPPLPLRESSPKKE 454

DB 705 ELPKGEQVVPDPAREAGPPE 727

RESULT 8

AA041452

ID AAM41452 standard; Protein; 699 AA.

AC AAM41452;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6383.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

XX Leukemia.

XX Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0468725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEO INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60608.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 such as central nervous system injuries -

Oy	430	-----EWYLCRWHQPPSPPLPUBRESSPKKKEIVNARCILMPSSVAGETSVLAAPSNRDH
Dd	842	DOEFHCNDNEFL-----EEA-----DELDADNPSELETIPKMDV 87/4
Oy	482	SVEPLRADPNPSDLLENLDDSVFSKRRAKLDELDERKKRWIDIQRIEQRILORLOLRMYK 54/1
Dd	875	PVPSSSDRIDTEL---LSDAFTERRRHQKVXKDEVRCKORDARVYMKEQJRLQLMRRNQD 93/3
Oy	542	KGIQESRP-EVTSFFPEPDVESMLITTPPLPVVAAGRPLPKLTPO-NFELPWLD 59/3
Dd	932	EVLALDPLRASFTFPLPEDIALIOFVNEEVTVQAEGENVVMNEARDPGVWVD 98/5
 RESULT 10 ABP41839 standard; Protein; 777 AA.		
ID	ABP41839	
XX	ABP41839;	
AC		
XX		
DT	22-AUG-2002	(first entry)
XX		
DE	Human ovarian antigen HNBVO53,	SEQ ID NO:2971.
XX		
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;	
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;	
KW	fertility; pregnancy disorder; anovulation; polycystic ovary syndrome;	
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;	
KW	inflammatory condition; immune disorder; blood disorder;	
KW	cardiovascular disorder; respiratory disorder; neurological disorder;	
KW	gastrointestinal disorder; urinary system disorder; drug screening;	
KW	gene therapy; chromosome mapping; forensic analysis;	
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;	
KW	antiinflammatory; gynaecological; reproductive.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200200677-A1.	
PD		
XX	03-JAN-2002.	
PF	07-JUN-2001; 2001WO-US18569.	
PR	07-JUN-2000; 2000US-209467P.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Blirse CE, Rosen CA;	
DR	WPI: 2002-147878/19.	
DR	N-PSDB: AB054916.	
XX		
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,	
PT	useful in the prevention, treatment and diagnosis of cancer (e.g.,	
PT	ovarian cancer), immune disorders, cardiovascular disorders and	
PT	neurological diseases -	
XX		
Claim 11:	SEQ ID NO 2971; 2922bp; English.	
CC	The invention relates to 2175 novel human ovarian antigens (ABP41054-	
CC	ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also	
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC	to the sequences of the invention. The invention additionally relates to	
CC	recombinant vectors and host cells comprising human ovarian antigen	
CC	polynucleotides, antibodies against human ovarian antigens, and the use	
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,	
CC	treating, prognosing or preventing various ovary and/or breast-related	
CC	disorders. Such conditions include ovarian cancer and breast cancer, and	
CC	metastatic tumours of ovarian or breast origin, reproductive system	
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,	
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine	
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic	
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and	
CC	vaginitis), immune disorders (e.g., congenital and acquired	

CC		immunodeficiencies, autoimmune ophthalmitis, systemic lupus erythematosus),
CC		respiratory disorders, neurological disorders, gastrointestinal disorders,
CC		and urinary system disorders. Ovarian antigen polypeptides and
CC		polynucleotides may also be used in screening for compounds which
CC		modulate ovarian antigen expression or activity. The polynucleotides may
CC		further be used for gene therapy, chromosome mapping, in the
CC		identification of individuals and in forensic analysis, and the
CC		polypeptides may be used as food additives or to prepare antibodies
CC		useful in disease diagnosis, drug targeting and phenotyping. The present
CC		sequence represents a human ovarian antigen of the invention.
CC		Note: The sequence data for this patent did not form part of the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences.
XX	SQ	Sequence 777 AA:
XX		
XX	Query Match	6.5%; Score 209.5; DB 23; Length 777;
XX	Best Local Similarity	23.1%; Pred. No. 1.1e-05;
XX	Matches 115; Conservative	51; Mismatches 204; Indels 127; Gaps 20
QY	55	GPPPLASSGG-----SPAPSGCGGKGRGLLPAGAAG--QQEBSWGGS 99
Db	347	PGPPTTAHNLGVSVGLVSPRLRPGENHRAGSNEDPILASGPPTTIPDEFFGR 406
QY	100	VPLPCPPRATQ-----AIGGERAAAGACGSRPKYQAVLPIQ 138
Db	407	VPRFAFYHDHEASDVLSLESDSDSVIVPEGLPLPPPPSGATPPP----- 457
QY	139	TGSLVAANKERTPMAGDKGAASPAAATASD-----PAGPPRLPGPPTLATATAGTLA 193
Db	458	----IAPGPP-----ASPPVPAKEEPDELPAABGPLP-FPPPPPVVPGPYILP 503
QY	194	ASEGHWKSMKRSPLGGGGSGASSOACLKOILLDLLEGOOOOLQAKEREIEELKSE 253
Db	504	PPO-----LVPECTPGGGPPALEDDLTVININSSD--EEEBEEEEEEEEBE 554
QY	254	KDTLLARIERNRKQLVKKNCKERNHLFGYTEEBERETLSKKIKLCQPELTSISQ 313
Db	555	ED-----FEEBEEDEEYEFEEBEEBEEFEFEFEGBELEBEEBEEDEEBEELEEV-- 607
QY	314	TLPPRFSCGSGSKHKRRKSPPGSTERTKYVKLAPERSKVTKTPKHSPIKEBOGSLS 373
Db	608	----EDLEFGTAG-----GEVEEGAPPPTPTLRPALPPESPYPKVOPEPEPERGILL 654
QY	374	E-----TVCKRELRSQ--ETPEKPSSVDTPPRLSTPOKGPSTHPKERAF 416
Db	655	EVEERGTEERGADATPLADEALPSQGEVEEREGSPAAPROPJELVEEPSAPP--TL 712
QY	417	SSEIDFLYLSTTEMYLTCRMHQPPSPPLRKSSSKKEETVARCLMPSSVAGETSVALVP 476
Db	713	EEETED-----GSDKV-----OPPE-TPAEER-METETEALOEKEDODTAAMLA-- 757
QY	477	SWRDHSVLEPLRDPNPSSD 493
Db	758	DFIDCPPDEKPPPETE 774
RESULT 11		
AAW31186		
ID	AAW31186	standard; Protein; 905 AA.
XX	AAW31186;	
XX	AC	
XX	DT	11-MAR-1998 (first entry)
XX	XX	
DE	Human p160 polypeptide 160.2.	
XX	p160; p62; cytoplasmic; T cell; B cell; development; activation;	
KW	modulation; cellular response; cell proliferation; autoimmune disease;	
KM	p56-lck.	
OS	Homo sapiens	

```

XX Key Location/Qualifiers
FH Domain 3..138
FT Region /label= leucine_zipper
FT 510..638
FT /note= "proline/lysine rich region"
FT 654..870
FT /note= "glutamic acid rich region"
XX
PN MO9722255-A1.
XX
XX 26-JUN-1997.
XX
XX 11-DEC-1996; 96WO-US19944.
XX
XX 19-DEC-1995; 95US-0574959.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Joung I, Shin J, Strominger JL, Vadiamudi RK;
XX
XX WPI: 1997-341351/31.
XX
XX N-PSDB; AAT89346.
XX
XX cDNA encoding p62 and p160 and corresponding proteins - used in the
XX treatment of autoimmune disease and for T and B cell proliferation,
XX e.g. for treatment of tumours
XX
XX Claim 84; Fig 11; 175pp; English.
XX
XX This sequence represents a novel p160 polypeptide (160.2) which is
XX capable of activating transcription of a variety of genes upon activation
XX of p62 and is capable of binding to the p62/p56lck complex to modulate
XX lck function in a manner similar to p62. The genes transcribed in
XX response to p160 activation likely include those of which are involved in
XX T or B cell development/differentiation, T or B cell activation or
XX production of T or B cell specific factors e.g. lymphokines or
XX antibodies. This p160 polypeptide is also a substrate for
XX serine/threonine kinase activity. p160 polypeptides can modulate
XX degradation of cellular proteins e.g. cell cycle regulatory proteins
XX stimulating expression of cell cycle dependent kinase inhibitors and
XX arresting cell cycle progression at specific boundaries to thereby
XX modulate cell proliferation. As p160 boosts B cell response it may be
XX used to treat disorders where this is beneficial, e.g. infections by
XX pathogenic microorganisms. p160 can be used to expand T cell populations
XX for treating infectious diseases or cancer and p160 inhibitors could
XX reduce B or T cell responses and may be used to treat a variety of
XX autoimmune diseases, e.g. diabetes mellitus, arthritis, multiple
XX sclerosis allergic reactions and Crohn's disease.
XX
XX Sequence 905 AA;
XX
XX Query Match 6.4%; Score 206.5; DB 18; Length 905;
XX Best Local Similarity 22.4%; Pred. No. 2.2e-05;
XX Matches 121; Conservative 52; Mismatches 231; Indels 137; Gaps 21;
XX
XX 38 PGAAEAHFPLRRHK-----LKEGPPLASSGGSPASPAGCGGGRGLLPAGAPG 90
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 414 PTPAPVPLRPHRSGFHRSLRAKPCQMA-----PCPQAPCPSSAG-----DMPSSGP 462
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 91 QOESWGG-----SVPLPCPP-----PATKAGTIGEPAAAGACS----- 126
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 463 VPSEPWTSTANLGLSRSVCPRLLPGENHRASSNEDPILAPSGTPTPTTPDET 522
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 127 -----PRPKY-----QAVLPQTGS-----LVAAKEPTWAGDKGAASPAAPASD 168
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 523 GGRVPRPAFYHDKESADVEISLESDDSVIVIPGGLPLPLPPPSGATPPPIATGP 582
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 169 PACPPPLPL-----PGPPLATPATAGTLAASERKMSMRSPISLGGGGGS 213
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 583 PTPASPPPAKEEPEELPAACGPPLPPPPPPPPVGVYLLPPQ-----LVPGTPEGG 635
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 214 GASSOAAQLKQILLQLDLIEQQOQOOLQAKKEIEELKSERDILLARIERMERMQLVK 273

```

```

Db 636 GPALREDLTFYININSSDEEEEGEHEEEEEEDEEEDEEE-----EEEDDFEEEE 690
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 DNE---KERKLPFGCYET-ERETELSEKIKLECOPELSTOTLPKPPSCGSGKG 329
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 691 DEEFYFEEHEEEHEEEHEEEHEEGELEEEHEEEDDEEEHEEEV---EDLEFGTAG-- 744
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 330 KRKSPFGSTERKTPVKKLAPFESKVKTKTPKHSPIKEEPCGSLSEYCKRELRSQETPEK 389
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 745 -----GEVEGAPPPPTLPALPPPSPPKVPQPEPEGLLEV-----EEPGTE 790
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 390 PRSSVDTPPRLSTPQKGPSTHPKKAFFSSEIEDLPYSTTEMYLCRWHPPPSPPLURE- 448
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 791 EERGADTAPTLA-DEALPQGEVREGESEPAAGPPQELVE-----EESSPPTLLEE 843
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 449 -----SSPKAE---ETVARCIMPSSVAGETSVLAVPSMDHSEVPLRDPNPS 492
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 844 TEDGSDRVQPPPEPFAEEMETETEALQEKEDDTAAMLA--DFIDCPDDEKPPPT 901
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 493 D 493
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 902 E 902
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
AAM31185
ID AAM31185 standard; Protein: 1135 AA.
XX
XX AC AAM31185;
XX
XX DT 11-MAR-1998 (first entry)
XX
XX DE Human p160 polypeptide 160.1.
XX
XX KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
XX modulation; cellular response; cell proliferation; autoimmune disease;
XX p56-lck.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
FH Domain 3..138
FT Region /label= leucine_zipper
FT 740..868
FT /note= "proline/lysine rich region"
FT 884..1100
FT /note= "glutamic acid rich region"
XX
XX PN MO9722255-A1.
XX
XX PD 26-JUN-1997.
XX
XX PF 11-DEC-1996; 96WO-US19944.
XX
XX PR 19-DEC-1995; 95US-0574959.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Joung I, Shin J, Strominger JL, Vadiamudi RK;
XX
XX WPI: 1997-341351/31.
XX
XX N-PSDB; AAT89345.
XX
XX cDNA encoding p62 and p160 and corresponding proteins - used in the
XX treatment of autoimmune disease and for T and B cell proliferation,
XX e.g. for treatment of tumours
XX
XX Claim 84; Fig 9; 175pp; English.
XX
XX This sequence represents a novel p160 polypeptide (160.1) which is
XX capable of activating transcription of a variety of genes upon activation
XX of p62 and is capable of binding to the p62/p56lck complex to modulate
XX lck function in a manner similar to p62. The genes transcribed in

```

CC response to p160 activation likely include those of which are involved in
 CC T or B cell development/differentiation, T or B cell activation or
 CC production of T or B cell specific factors e.g. lymphokines or
 CC antibodies. This p160 polypeptide is also a substrate for
 CC serine/threonine kinase activity. p160 polypeptides can modulate
 CC degradation of cellular proteins e.g. cell cycle regulatory proteins
 CC stimulating expression of cell cycle dependent kinase inhibitors and
 CC arresting cell cycle progression at specific boundaries to thereby
 CC modulate cell proliferation. As p160 boosts B cell response it may be
 CC used to treat disorders where this is beneficial, e.g. infections by
 CC pathogenic microorganisms. p160 can be used to expand T cell populations
 CC for treating infectious diseases or cancer and p160 inhibitors could
 CC reduce B or T cell responses and may be used to treat a variety of
 CC autoimmune diseases, e.g. diabetes mellitus, arthritis, multiple
 CC sclerosis allergic reactions and Crohn's disease.

XX Sequence 1135 AA:

Query Match 6.4%; Score 206.5; DB 18; Length 1135;
 Best Local Similarity 22.4%; Pred. No. 2,9e-05;
 Matches 121; Conservative 52; Mismatches 231; Indels 137; Gaps 21;

QY 38 PGAAEAHFLPRHKK-----LKEPPPLASGSGSPAPSPAGCGGKRGRLLPAGAPG 90
 DB 644 PTPAPPPPLRPHRSPGPHSLTAPCPDMA-----PCPQACPPSAG-----PMSAGP 692

QY 91 QQESWGG-----SVPLPCPP-----PATKAGIGGEPAAAGAGCS----- 126

DB 693 VPSEPTSTTANLGLLSRPSVCPRLPPEHNHRAAGSNDPILAVSGTPPTIPDET 752

QY 127 -----PRPKY-----QAVLPICGGS-----LVAAKPTPWAGDKGAAPATASD 168

DB 753 GGNVPRAAFVHYDKESADVEISLESDDSVIVIEGCLPRLPPSPSGATPPPIATGP 812

QY 169 PAPPPLPL-----PCPPPLAPTATAGTLAASGRMKSMKSPGLGGGGS 213

DB 813 PTASPPVPKAEPEELPAPAGPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 865

QY 214 GASSQAACLKQILLDLLEQQOQOLQAKKEIEELKSRDYLARIEMERRMOLYKR 273

DB 866 GPALDEEDLVININSDEEEEGEEDDEEEEDDEEEEDDEEEEDDEEEEDDEEEEDDEEE 920

QY 274 DNE---KERHKLQGYETE-ERETELSEKIKLECGELSETSOTLPKFFSGSGSGKH 329

DB 921 DEEEYEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 974

QY 330 KRKSPGSGTFRKTPVKLAPEFSKVTKTPKHSPIKEEPCGSLSEYCKRELISOETPEK 389

DB 975 -----GEVEGAPPPPTLPALPPSPKVPPEPPPPGGLLEV-----EEBGT 1020

QY 390 PRSSVDTPPRLSTPKGPTSHPKAFSSIEDLPYLSTTEMLCRKHQPPSPPLPLRE- 448

DB 1021 EERGADTAPTLA-PEALPISQGEVEREGESPAAGPPPELVE-----EERSPPPTLLEE 1073

QY 449 -----SSPKKE-----ETVARCLMPSSVAGSTVLAVPMSRDHVEVLKDPNPS 492

DB 1074 TEDGSKVOPPPPTPAEEMETETEAALQEKQDDTAAMLAA--DFIDCPDDEKPPPT 1131

QY 493 D 493

DB 1132 E 1132

RESULT 13
 AA082954 standard; protein; 1023 AA.

AA082954;

23-APR-2002 (first entry)

Human homologue of MPT1 protein target for antifungal compound.

Human homologue of MPT1 protein target for antifungal compound.

Human homologue of MPT1 protein target for antifungal compound.

KW Antifungal; fungal gene transcription; RPC34; POB3; TFA2; NAB2;
 KW MPT1; MTR2; BOS1; POL30; RSA2; SOT1; MTW1; SPC98; BFR2; RNA1;
 KW GCD7; SK16; NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4;
 KW yeast; fungus.

OS Homo sapiens.

PN WO200202055-A2.

PD 10-JAN-2002.

XX 28-JUN-2001; 2001MO-US20592.

XX 29-JUN-2000; 2000US-215164P.

PR 10-AUG-2000; 2000US-224457P.

XX (ANAD-) ANADYS PHARM INC.

XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S, Long F;
 PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S,
 PI Davidov E, Thompson CM;

DR WPT: 2002-147962/19.

XX N-PSDB; ABK32842.

PT Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects

PS Claim 1; Figure 79; 522pp; English.

XX The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g. fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POB3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
 CC SOT1, MTW1, TFB1, SPC98, BFR2, RNA1, GCD7, SK16, NIP1, LCP5, NCE103,
 CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly C. albicans growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This amino acid sequence represents a
 CC target protein used to test the antifungal compounds, described in the
 CC method of the invention.

XX Sequence 1023 AA:

Query Match 6.4%; Score 206; DB 23; Length 1023;
 Best Local Similarity 22.1%; Pred. No. 2,7e-05;
 Matches 152; Conservative 72; Mismatches 264; Indels 200; Gaps 28;

QY 11 AAAPA-----GNGEORLDERAALGPEDEPGAABHFLPRHRLKKEPPPLASSGGS 66

DB 208 AAAPAVSLVNNGPALPLPCKPAPGTVIOTPFVGAAPAPAPAPAPAPAPAPAPAA 267

QY 67 PAPPSPAGCGGKRGRLLPAG--AAB-----GQCESWGSVPLPCPPATKQAGIGE 117

DB 268 PPPPPAPATIANRPHGPPAPPAAPVPPPAAGNGSAGAPAPAP-PAAGSPAGVSQ 326

QY 118 PAAAGAGCSBPBKRYQAVLPIOTGSLVAAAKEPTPWAGDKGAASPAATAS----- 167

DB 327 PCGGAATAAPADPCKAESPKR--VVOAA-----PPAAQTILASGPASTAASWICPTMG 379

QY 168 -----DPAGPPPLPPLPAGPPPLAPTATAGTLAASGRMKSMKSPGLGGGSGASSQAACLR- 223

DB 380 ALPSPAAPP-PAAGTPGTGLPKGAAGAVT-----QSLSRTPATTTGIRATILPTVLAP 432

```

OY 224 -----OILLQLDLLEQOOQLOAKK----- 245
DB 433 RLPPPPNPTNIQNFOLPRGAVLYRSENGOULMIPQALMOAOAHQAPOTTPMAPPAT 492
OY 246 -----EIEELKSERDTLLAR-----IERMERMOVLKDKNEKER----- 279
DB 493 PLSAPVQISTVQAAPRIIARQVTPPTIIKQVQAQOTTVQPSATLQSPGVQPOLVLGG 552
OY 280 -----HKLFQGYETEEREETELSEKIKLECOPELSETSQTLPPKPF 321
DB 553 AAGTASLGATAVQGTGPQRTVPGATTTSSATETMENVK-KCKNFLS-----TL-----IK 603
OY 322 CGRSGKHKKRSPGSTERKTPVKKL-----APEFSKVTKTPKHSPIKEEPCGS 371
DB 604 LASSGK-----QSTETAAVKELVONLIDKIEADFTSRLYRELNSSP---QP--- 649
OY 372 LSETVCKREL-----RSQETPEKPRSSVDTPPRL-----STFOKGPSTHPK 412
DB 650 YLVPEFLKRLPALRQLPFDAAFIQSQSQQPPPTSQATTLALVVLSSVQR---TAGK 706
OY 413 EKASSSEEDLPYISTE-MYLCRMHQPPSPPLPRESPPKKEETVARCLMPSSVAGETS 471
DB 707 TAAVTYSALQPPVLSTQPTQVGVKQGOQPPPLVIOQ--PPKPGALIR--PPQVTLIQTP 762
OY 472 VLAVPNRDHSVEPLRDPNPSDLLENLDDSVFSKRHAKL-----ELDEKRRKMDIQIR 526
DB 763 MVA-----LROPNRIMLTTPQOVNLSEBSARILATNSELVGTTRSCDKETFL 811
OY 527 EORILQRLQRLMYRKKGIOESEPEVTSF 554
DB 812 LQAPLQRRILEIGKKHGITELHPDVVS 839

```

RESULT 14

AAM39666 standard; Protein: 676 AA.

AAM39666;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2811.

Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;

peripheral nervous system; neuropathy; central nervous system; CNS;

Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

chemokine; thrombolytic; drug screening; arthritis; inflammation;

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

leukaemia.

```

XX WPI: 2001-442253/47.
DR N-PSDB: AA158822.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS Example 4: SEQ ID NO 2811; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA161365) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
SO Sequence 676 AA:

```

```

Query Match 6.3%; Score 205.5; DB 22; Length 676;
Best Local Similarity 22.0%; Pred. No. 1.8e-05;
Matches 131; Conservative 82; Mismatches 224; Indels 159; Gaps 25;

```

```

OY 55 PGPPLASQGGSPAPRAGC---GKGGILLPAGAGQOEEMSGSVPLPCPPRATKO 111
DB 2 PSAKQKSGKGGHGAASPEKGAHPSAARPLAPRAACSPSPGGA---PASPPGAP 58
OY 112 AGIGEPAAAGAGCSPRKYQAVLPIOTGSLVAAKKEPTPWAGDKG--GAASPAATSDP 169
DB 59 RSLASQPARAAARAPAM-----SAKQ---RSGKGGHGAASPEKGAH 99
OY 170 AG-----PPPLPGLPPLA-----PPLATGTLAASGGRMKSMKSPSLGGGGSG- 214
DB 100 SGADADVAKKPPRPAQOPPPRPAHPOOHPOHNGKGRH-----GGGGGGGK 152
OY 215 -----ASSQAACIQI-----ILLQLD-----IEQOOQLOAKE 244
DB 153 SSSSSASAAAAAASSSASCSSRLGRALNLFYLAUAAAFSGWCVNHVLEVOVR 212
OY 245 KEIEELKSERDTLLARIMERRMOLVKDKNEKERHKLFGQYETEEREETELSEKIKLDC 304
DB 213 RSHQDFSKQREBELQGLGVQKQSL-----QATFGFESILSSQHKODLTAAYVKG 267
OY 305 QPELSETSQTLPP-----KPFSGRSGKHKRKSPEGS-----TERKTPVKKLAP- 350
DB 268 ESEVSRISSEVLQKLONELIKDLSQIHVKKDAREDFSLNTEVERTELTKSINDNIA 327
OY 351 -FSKYKTKTPKHSPIKEPCGSLST-----VCKRELQSQETPEKPRSSVTPPRLS 401
DB 328 IFTEVQKRSQKEINDMKAKVASLSESEGNKODLAKLAKVAKEIQTSASR-EMDEALRS 386
OY 402 TPQ-----KGPSTHPKEKAF--SSEIEDLPYSTEMYLCRMHQPPSPPLURE 448
DB 387 TLQTMESDLYTEVRRLVSLKQEQQFKRAADTERLALDALTEKLL----- 431
OY 449 SSPKKEETVARCLMPSSVAGETSVIAPVSMNDH-----SVEPLDPPND--SDL 494
DB 432 ---RSEESVSR--LPEEIRRLLEEELRQLKSDSHGCKEDGGRHSEAFAPALOOKSGDLDSR 486
OY 495 LENLDDSVFSKRHAKLELDEKRRKMDIQIRREORILLQRLQRLMYKKKGIOESEPE 550
DB 487 LQHVDEGLVSMQVASAROTESLESLSKSQEHQR-LALQGRL---EGLSSSEAD 538

```

(HXSE-) HXSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Zhang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

RESULT 15
 ID AAY85565 standard; Protein; 2432 AA.
 XX AAY85565;
 XX 07-JUL-2000 (first entry)
 XX Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
 DE UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
 KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 KW antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 1..89
 FT /note= "this region can be replaced with one of the
 FT three sequences shown in AAY85566 to AAY85568;
 FT this creates three variants at the N-terminus"
 FT Misc-difference 1018
 FT /label= Asp or Glu
 FT 1776..1778
 FT /note= "present or absent depending upon the allele
 FT from which the protein is translated"
 FT Region
 XX
 XX MO9963080-A1.
 XX
 XX 09-DEC-1999.
 XX
 XX 02-JUN-1999; 99WO-EP03848.
 XX
 XX 03-JUN-1998; 98GB-0011962.
 XX
 XX (JANC) JANSSEN PHARM NV.
 PA
 PI Luyten WHL, De Raemaeker MC, Geysen JGH, Bogaert TAOC;
 PI Maerten LJS, Verhasselt P, Van De Craen M;
 XX
 XX WPI: 2000-116370/10.
 DR N-PSDB; AAA07836.
 DR
 XX Novel proteins and nucleic acids e.g. for treating neurodegeneration
 PT
 XX
 XX Claim 95; Fig 1d; 146pp; English.
 PS
 XX The invention provides vertebrate (human) protein homologue of a UNC-53
 CC protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
 CC their plus ends. The UNC-53 sequences are used to promote neural
 CC regeneration, revascularization and wound healing; also for treating
 CC neurodegenerative disease, acute traumatic injury, fibrotic disease and
 CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
 CC polynucleotides can be used for recombinant production of the proteins,
 CC as a source of probes for detecting allelic variants and polymorphisms,
 CC for sequencing genomic DNA and for detecting UNC-53 expression; and as
 CC source of therapeutic antisense sequences. Cells that express the
 CC protein are used to identify regulators of cell shape, growth, motility
 CC and migration. They can also be used to identify proteins that are
 CC involved in signal transduction pathways also involving UNC-53, and to
 CC identify compounds that alter attachment of UNC-53 to microtubules. A
 CC target gene coupled to a UNC-53 encoding sequence may be used to deliver
 CC the target gene to a cellular microtubule or its plus ends. The present
 CC sequence represents the amino acid sequence of the second human homologue
 CC of UNC-53, designated hs-UNC-53/2.
 CC
 XX
 XX Sequence 2432 AA:
 SQ
 Query Match 6.3%; Score 203.5; DB 21; Length 2432;
 Best Local Similarity 22.3%; Pred. No. 0.00012;
 Matches 130; Conservative 81; Mismatches 259; Indels 113; Gaps 21;
 QY 6 AVTKAAAPA---CGNPEQLDVERAAALGPEDEPGAALAHFLPRHRKLKEPPLASS 62

Db 209 AVSQVAGAPSCQAGTPOQOV-----PVPQAPCCPHQAPAHQSKAEMQSRL 238
 QY 63 QGSPAPAPAGCGCGKGRLLLPAGAAPGQOESEWGSVPL--PCPPATKQAGIGGPAA 120
 Db 259 SGPTARVSAAGSEAKTGCGTTANNRRSOSFNNDKSKPTVSPPPSSHEK-----EPLA 314
 QY 121 AGAGCSPPPKYQAVLPQTGSLVAAKKEPPPMAGDKGAA-----SPAATAPDPAPPL 175
 Db 315 SSASSHPGMSDNAPASLESSESSPTNCSTSAIPQGAATKPMRSKSLVKSATVSM 374
 QY 176 PLPEPPPLAPATAGTTLAASBGRKSMRK-----SPLGGGGSGASQAACLKQILLIQ 229
 Db 375 SVKPPGEAPRPPEAKKPAAPNOKSMLEKTLKFNKSGSKAGGPGSPRDTSCERETLP 434
 QY 230 LDLIEQOQOOLQAKEKEIEEL---KSERDTLARIEMRMRLVKD---NEKERHK 281
 Db 435 ---SFESELEAASRMILTVGPASSSPKIALKGIORTSRALTNKSSLKGNKEKEK 491
 QY 282 LFGYTEEREER---TELSEKIKLEQPELSETSQTLPPKPSGCRSG---KGHK-- 330
 Db 492 QOREKDKKSKDLAKRASVTERLDLKEPKEDPSGAAVPMPKSKRIASFIRPGKILNS 551
 QY 331 -RKSPFESTERKTPVKKLAPBFSVKYKTKRK-HSPIKEEPCGSLSETVCRELRSQTP 388
 Db 552 AKKEPMAPSHSGIP---KQGMKSMGKSPSAPAPSK-----GERSRSGKLSSGIPQ 601
 QY 389 KP-----RSSVDTPPRLSTPQKGPETHPKERKAFSE-----IEDLPYLSTENYLCRMH 437
 Db 602 KPOLDGHHSSSSSLASSEKGRGCTILNHSISSQIVSGVGTQTGGSTIVSYOL----- 657
 QY 438 QPPSPPLPRESSPKKEETVARCIMPSSVAGETSVLAVPMSRDHSEVPL----- 486
 Db 658 ----PQPOQOYHNPATVAPFLYRSQDTDEGNVTESSSTGVSEPSHFTKTGPAL 712
 QY 487 ----RDP-----NPSDLENTLDSVSKR-----HAKLE 511
 Db 713 ELTGEDPEARRLRTVKNIAOLRQMLEETMSSLRGTOVTHSTLE 755

Search completed: July 14, 2003, 11:48:31
 Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 14, 2003, 13:41:51 ; Search time 4488 Seconds

(without alignments)
3981.534 Million cell updates/sec

Title: US-10-054-935-2
3238
Sequence: 1 MTMRSAVFKAAPAGGNPE.....RSRCRLRIQKQPHRTCRK 614

Scoring table: BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n model -DEV-xlh
-Q/cgn2_1/USPRO.spool/US10054935/runat_14072003.115025_24625/app.query.fasta.1.775
-DB-genemb1 -QPM-fastap -SUFFIX-rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human4.0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054935.ecgn.1.1.2819.0runat.14072003.115025.24625 -MCPU=6 -ICPU=3
-NMAP -LARGOQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vt:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1757.5	54.3	196724	10 AL590963	AL590963 Mouse DNA
2	1460	45.1	180017	2 AC068669	AC068669 Homo sapi
3	1311	40.5	193189	2 AC119462	AC119462 Rattus no
4	1055.5	32.6	615	6 AX401049	AX401049 Sequence
C 5	944.5	29.2	174346	2 AC131356	AC131356 Rattus no
C 6	906.5	28.0	193189	2 AC119462	AC119462 Rattus no
7	680.5	21.0	1433	9 HSM800235	AL049450 Homo sapi
8	416	12.8	207214	2 AC107643	AC107643 Mus muscu
9	380.5	11.8	174346	2 AC131356	AC131356 Rattus no
10	299	9.2	144577	2 AC123367	AC123367 Rattus no
11	274.5	8.5	207214	2 AC107643	AC107643 Mus muscu
12	254	7.8	136755	2 CNS08CB6	ALB45344 Oryza sat
13	246.5	7.6	3217	5 AB026623	AB026623 Gallus ga
14	246	7.6	3385	9 AB020652	AB020652 Homo sapi
15	246	7.6	3774	9 AF203032	AF203032 Homo sapi
16	246	7.6	5983	10 AF076623	AF076623 Mus muscu
C 17	243.5	7.5	16832	10 AB070943	AB070943 Streptomy
18	240	7.4	7033	9 AF039571	AF039571 Homo sapi
19	238	7.4	2313	9 BC014185	BC014185 Homo sapi
20	238	7.4	9551	6 AR076233	AR076233 Sequence
21	238	7.4	9551	6 I39845	I39845 Sequence 93
22	238	7.4	9551	9 HMMTRHYAL	L09190 Human trich
23	237	7.3	20293	1 SCD20	AL392148 Streptomy
24	236.5	7.3	1222	5 AF239885	AF239885 Gallus ga
25	236	7.3	7033	6 AX010193	AX010193 Sequence
26	234	7.2	2167	9 AB071115	AB071115 Macaca fa
27	234	7.2	8469	9 HSA7041	AJ007041 Homo sapi
28	234	7.2	177607	2 AC119484	AC119484 Rattus no
29	234	7.2	303976	2 LMFICHR15	AL160371 Leishmani
30	233.5	7.2	2800	9 AB040976	AB040976 Homo sapi
31	232.5	7.2	5858	3 AF350276	AF350276 Hemipila m
32	231	7.1	150366	2 AC119064	AC119064 Sus scrofa
C 33	231	7.1	151578	9 AL589986	AL589986 Human DNA
C 34	230.5	7.1	29641	3 AC005802	AC005802 Leishmani
35	230.5	7.1	110000	3 AC125735_0	AC125735 Leishmani
36	230	7.1	22970	9 HSL247F6	Z68279 Human DNA s
37	229.5	7.1	3621	9 EC005847	EC005847 Homo sapi
38	229.5	7.1	139517	9 AC105446	AC105446 Homo sapi
C 39	228	7.0	10091	1 AF135398	AF135398 Therinus b
C 40	228	7.0	146506	8 OSJN00206	AL663007 Oryza sat
41	227.5	7.0	8459	9 AB006626	AB006626 Homo sapi
C 42	227.5	7.0	22693	9 HSPLEB1S3	U63610 Human plect
43	227	7.0	177877	9 AL450472	AL450472 Human DNA
44	226.5	7.0	15534	9 AB051895	AB051895 Homo sapi
45	226	7.0	9121	4 OCCTRICHA	Z19092 O. cuniculus

RESULT 1

ALIGNMENTS

AL590963/c
LOCUS AL590963 196724 bp DNA linear ROD 05-APR-2002
DEFINITION Mouse DNA sequence from clone RP23-395E10 on chromosome 11,
complete sequence.
ACCESSION AL590963
VERSION AL590963.11 GI:20068458
KEYWORDS HTG,
SOURCE house mouse,
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Lovell, J.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced GI:16304739.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-395E10 is
from the RPCR-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBac3.6.
FEATURES
source location/Qualifiers
1..196724
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-395E10"
/clone_1lb="RPCI-23"
BASE COUNT 49842 a 49408 c 48640 g 48834 t
ORIGIN
Alignment Scores:
Pred. No.: 3.53e-43 Length: 196724
Score: 1757.50 Matches: 440
Percent Similarity: 29.53% Conservativeness: 8
Best Local Similarity: 29.00% Mismatches: 10
Query Match: 54.28% Indels: 1059
DB: 10 Gaps: 4
US-10-054-935-2 (1-614) x AL590963 (1-196724)
QY 1 MetThmetarGserAlaValPheLysAlaAlaAlaProAlaGlyLysAnProGlu 20
Db 62746 ATGACCATGAGATCCGACAGTGTTCAGGGCGCGCGCCCTCCGCGCAACCCCGAG 62687
QY 21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
Db 62686 CACGCACTGACACTACGACGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62627
QY 41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAla 60
Db 62626 GCGGAGGCCATTTCGCCCCCGGCACTGTAAGCTCAAGAGGCGGCGCGCGCGCGCG 62567
QY 61 SerSerGlnGlySerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79

|||||
Db 62507 TCTTCCAGGGCGGAGCCCTCGCTCCAGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 62507
QY 80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnLysSerPylGlySer 99
Db 62506 GGCTTTGACTCCAGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62447
QY 100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyLysGlyGluProAla 119
Db 62446 GTGCCCTTGCCCTGTCGCGCCCGCAGCTACAAACAGCGGCATTCGCGGGAGCCAGTC 62387
QY 120 AlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProLysGlnThr 139
Db 62386 GCAGCGGCGCTGCTGCACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62327
QY 140 GlySerLeu---ValAlaAlaAlaLysGluProThrProThrPylAlaLysPylGlyLys 158
Db 62326 GGCTCTATCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62267
QY 159 AlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuPro 178
Db 62266 GCGGCTGCCCTGACCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62207
QY 179 GlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlyArg 198
Db 62206 GGGCCACACACCTCG 62147
QY 199 TrpLysSerMetArgLysSerProLeuGlyGlyGlyLysSerGlnLysSerGln 218
Db 62146 TGGAGAGATTAAG 62087
QY 219 AlaAlaCysLeuLysGlnLeuLeuLeuGlnLeuAspLeuLysGlnGlnGlnGln 238
Db 62086 GCGGCTGCTCAACAGATCTTGTCTGCTCAATGAGACTCATCAACAGCAGCAGCAG 62027
QY 239 GlnLeuGlnAlaLysGlnLysGlnLysGlnLysGlnLysSerLysArgAspThr 256
Db 62026 CAGTTCGAGGCCAAG 61967
QY 256 ----- 256
Db 61966 GAGGGGTATATCTGCATTAGGAGATGGAGAGACGACTCTGCGCATGCTCGGGGAGAGGGG 61907
QY 256 ----- 256
Db 61906 CTGTAGAGAGCAGAGGTCCCGCTCGGTAGGTTAGAGAGAGTGGCCACGCTAAG 61847
QY 256 ----- 256
Db 61846 AGTGCTTAGATGTGGGTCTTAGAGAGCCTGACATCACTGACACGTTGGGTGGAAGA 61787
QY 256 ----- 256
Db 61786 AGGGTTAAAGGCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 61727
QY 256 ----- 256
Db 61726 AAAGGTTAATTTAAACAGACAGCCACACTGTGGGAAAGAGAGCTTTAGAGTGAAG 61667
QY 256 ----- 256
Db 61666 ACCTGATTTAGAGGGGTTAAAAATGATAGAGGTGGGGTGCAATGTGGCTGATGACAG 61607
QY 256 ----- 256
Db 61606 GGTGTATTCACAGCGCTGTGCTTAGCTGTGATTCGATTGAGAGAGAGGCTATACAGCTGTT 61547
QY 256 ----- 256
Db 61546 GCGGTAACACCGCTGTGTGATCTTTAGACAGAGAGGTTGTCTATAGTCTTTTGTGCGGT 61487
QY 256 ----- 256

Db 61486 AGAGGTAAAGATGGTGTACGTGGAATCCTGTTTACCTCTCTAAATAGAAATCACTGC 61427
QY 256 ----- 256
Db 61426 CTTTGATTTAGAGAGTTTCTAGTAAATCTGATTAATAAACTACCCGCTGACCA 61367
QY 256 ----- 256
Db 61366 CGTCCCTTGAAGTGGCAGATTTTGAGACATAAGCTTATTCGATCTATTCTGTAAAC 61307
QY 256 ----- 256
Db 61306 CCTTCACCTCCAGACCTGTTTTTGAATCTTGAGTATTAGAAAAGTTGTTGTTATTT 61247
QY 256 ----- 256
Db 61246 CTTTGTACTCTGGAAAGCATGACTCATAGGCTTGTCTGATGAGCAGACGCTTGGGC 61187
QY 256 ----- 256
Db 61186 ACACATGTTAAGCACACCTCCCTCACACCTCAGGGCCGTGGGGTGGTGGTGTACAGTG 61127
QY 256 ----- 256
Db 61126 GTAAATACCTTTTGAAGCGCTAAATAAATACATGGCTTACAGTGGCTCAGGCCCTG 61067
QY 256 ----- 256
Db 61066 CTCCCCCTCCCTCCCTGCTGGCTCTTACTCTGGGGCTTCACGACACAGGCTGAG 61007
QY 256 ----- 256
Db 61006 CATGGGAGACAAGAGCTTTCATTTTGTTCCTTAATAGTGGCATCAGGCCCTC 60947
QY 256 ----- 256
Db 60946 TGTATTGATGAGAGTGCCTTCTGCAAGTGAACAGAGAGCTGGGGCAGCTTGTGAA 60887
QY 256 ----- 256
Db 60886 ATACCTCAGTGTCTTCTTGTGGAAATACCTCAGATGAGTGGACCTGCTCAGCAGCTGA 60827
QY 256 ----- 256
Db 60826 CTGAGTCTACGATGTGATGCGCCCGCTAAACACACATATCCTCTTCATGGG 60767
QY 256 ----- 256
Db 60766 AAGGGGTGGAGAGAAAGATCAGTAAGGAGAAATGAGACTATGTACCCACGA 60707
QY 256 ----- 256
Db 60706 AACATGGCTAAGTAATGCTAAGGCTGTGAATCAGAACACAGGTGACACGAGTGGCTGCA 60647
QY 256 ----- 256
Db 60646 GGGGGGGGGGACACTACTGCACTGAGAGGTTGTCTGAGTTCTTAGGTGATGCCCTT 60587
QY 256 ----- 256
Db 60586 CTTTGGCTCCTTCATAGGCTTTTCAMAGTAGGAAAGATGGGGTGAATATGCTAAG 60527
QY 256 ----- 256
Db 60526 CATTTGTCTAGCTTTGGCATTACAGGCTTCTGTCTACTACTAGCTTTAGCCCTGCTTA 60467
QY 256 ----- 256
Db 60466 GGCTGCTTATTTCTAGTACTTGTGGCTCTGCTGAAGTCAATGCCATGGCAGTAG 60407
QY 256 ----- 256
Db 60406 AGACAGCAGTGTGACCTACTACTAACCAAGCAAAACTCTTGTCTTGTCTATTAAGCAG 60347

QY 256 ----- 256
Db 60346 CACAGTACCTTTTGTCAAAACTGAGTTAATACCAAGTTTGTGATGACATACCTTCC 60287
QY 256 ----- 256
Db 60286 ATTTTGTGATTTTGTGTTGCTCTTAAATAGATGAATTTTGAATAATATTAGAT 60227
QY 256 ----- 256
Db 60226 AATATTTTCCGTATCTCCACACAAAGCAGAACTATATATGTCATGTGTGGAT 60167
QY 256 ----- 256
Db 60166 TATATTCTGGCTGATTAATATTCTTTTCAGTGTAGTATGTTTGAATTACTCCCTC 60107
QY 256 ----- 256
Db 60106 AAATACGAATCATACATTTTGAATGTGTACTGTCTATCCCTAGGAGGCTGTGAGCC 60047
QY 256 ----- 256
Db 60046 CTTGATTTCTGATGTGTGCTTTTAAACATTCACGTGTCTAGATTGAGTGAA 59987
QY 256 ----- 256
Db 59986 CATAGACAGTATACATTTTGACACACCGTAGACACCTTAAAGACACAGGCTGAGAA 59927
QY 256 ----- 256
Db 59926 TGTAGTCAAGTGTGTGATGCTTAGAGTGTTCAGAGCAGGCTCAGAGGCGCTGATT 59867
QY 256 ----- 256
Db 59866 CCATCCCTAGCAGACAGACGACGACACGACAAAGAAAGATTGAGAGGGTCC 59807
QY 256 ----- 256
Db 59806 CAGAAAGCTTGAGACCAAGTGGCAGAGGCTTGCCATTAAAGGACATCCGCTGTGCC 59747
QY 256 ----- 256
Db 59746 AGTGGAGAGGCTCCCTTCAGAAAGAAAGACTTCAGGCTGCCAGTGGAGAGGCTCC 59687
QY 256 ----- 256
Db 59686 TTCAGAAAGAGAGCTTCAGGCTGCCCTAGCTCTTGCCATCAGACCAACTCCTACAT 59627
QY 256 ----- 256
Db 59626 TACATTTAGTGTCTTTGCTTCATTGGGTTTAAAGGAAGCTTTCAGATTGCTGCTG 59567
QY 256 ----- 256
Db 59566 CCAGTTCAGAGTAAAGTTGAATTGTAGGAGCATACGCGACGACAGAAAGATGCTGAT 59507
QY 256 ----- 256
Db 59506 TCCATAATGGGCTAAGTACGCGGCTTAACTTGTGTGTAGTCTGTACCTTGTCACT 59447
QY 257 ----- 257
Db 59446 TCCCTCCAGCTCTCTGCTCGATTTGAAGTATGGAAGGCGGATGCACTGTGTAAGAG 59387
QY 274 AspAsnGluysGluAtyGhisLysLeuPheGlnGlyTrpGluThrGluArgGlu 293
Db 59386 GATAACGAGAGGAAGGACACACTGCTCCAGGAGCTATGAACCTGAAGAGAGAGGAA 59327
QY 294 ThrGluLeuSerGluysIleLysLeuGlnLysGlnProGluLeuSerGluThrSerGln 313
Db 59326 GCAGAGTTGCTGTGAAATAATTAAATTGTGAGGCGACGCGGAGCTTTCGAGAGACTCCAG 59267

QY 314 ThrLeuProProlysProPheSerCysGlyArgSerGlyLysGlyHisLys----- 330
 Db 59266 GCTTCGCTTCCAAAGCCTTTCTCATGTGCGAGTGGAAAGGACACAAAGAGTGTGCT 59207
 QY 330 ----- 330
 Db 59206 GGTCACTTATTTCTTACTGAAAAACATGAGAGCTGCTAATGTGATGATATATGAG 59147
 QY 330 ----- 330
 Db 59146 CTTGGGGTTCGTGATGGATTAATAGCAACCAATACCTGATGAGAAGTCTGTTATTC 59087
 QY 330 ----- 330
 Db 59086 TAGGGGATGCTGCTTACCCCTGTAATACAGACACTTAGATGCTAAGCAAGATCT 59027
 QY 330 ----- 330
 Db 59026 GTTGTGAGTTCAAGTCAGCTGCGACAGTAGGAACCTGTTTTTACAAACAAAGCTTC 58967
 QY 330 ----- 330
 Db 58966 CAGTTTCTGTGTGAGAAATGCTTGACCACTTTCTGTGTCACAGTACTAGTACTC 58907
 QY 330 ----- 330
 Db 58906 GGGTGCACGTCACCTCCCAAGAGCCCTGCTGTATCTTGAAAGCTGACCCCTTGCTGC 58847
 QY 330 ----- 330
 Db 58846 CCATGGCTTCCAAAGCAAGACAGTATGATTTAGTACTGTAACCTCTCTTGCCCTCAA 58787
 QY 330 ----- 330
 Db 58786 TATATTCAAATCAATTTTATTTGAAATCAATTAATTTTGTGTTTGTGTTTCG 58727
 QY 330 ----- 330
 Db 58726 AGACGGGGTTCCTCTGTGTAGACCAAGCTGCGCTCGAAGTCAAGAAATCTGCGCTGCGC 58667
 QY 330 ----- 330
 Db 58666 TCCCTAGTGTGGATTAAGGCGGTGTCACCACTGCGCTGCGCTGAAATCAATTAAT 58607
 QY 331 -----ArgLysSerProPheGlySerThiGluArgLys 341
 Db 58606 TTTATTAACATTTTCTTCTTACCTAATAGGAAACCCCATTTGGAAATACAGAAAGAAAG 58547
 QY 342 ThrProValLysLysLeuAlaProGluPheSerLysValLysThiLysThiProLysHis 361
 Db 58546 ACTCCTGTTAAAAAGCTGCTGCTGCAATTTTCAAAAAGTCAAAAACAAACTCTTAGGCAC 58487
 QY 362 SerProIleLysGluGluProCysGlySerLeuSerGluThrValCysLysArgLysLeu 381
 Db 58486 TCTCCCATTAAGAGGAACCTGTGTGCTCATATCAGAAACTGTTTGAACGGAATG 58427
 QY 382 ArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLeuSer 401
 Db 58426 AGGAGCCAAAGAAACCCAGAAACCCCGGCTTCAGTGTATCCCAAGCAACTCTGCG 58367
 QY 402 ThrProGluLysGlyProSerThrHisProLysLysAlaPheSerSerGluIleGlu 421
 Db 58366 ACTCCCAAAAGGAGACCCAGACCCACCCCAAGAGAAAGCTTCTCAAGTGAATGGAA 58307
 QY 422 AspLeuProTyrLeuSerThrThiGluMetLysCysArgTyrPheGlnProProPro 441
 Db 58306 GATTTGCCGTACCTTTCACACAGAAATGATTTGTGTGCTTGACACACACCTCCCA 58247
 QY 442 SerProLeuProLeuArgLysSerSerProLysLysGluGluThrValAla 458
 Db 58246 TCACCGCTTACCATTCAGGGAATCTCTCCAAAGAGAGAGACTGTAGCA 58196

RESULT 2

AC068669 180017 bp DNA linear HTG 16-FEB-2002
 LOCUS Homo sapiens chromosome 17 clone RP11-749116 map 17, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 16 unordered pieces.
 AC068669
 AC068669.4 GI:18677458
 HTG: HTGS_PHASE1; HTGS_FUILLTOP; HTGS_ACTIVEFIN.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 180017)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 17, clone RP11-749116
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 180017)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Lacroque, R., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McCurt, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Menes, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Tridillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 180017)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
 Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
 MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrim, J., Menes, L., Mihova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Tridillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Feb 15, 2002 this sequence version replaced g1:11597011.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L10045
Center clone name: 749_L16

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 7782: contig of 7782 bp in length
* 7783 7882: gap of 100 bp
* 7883 15614: contig of 7732 bp in length
* 15615 15714: gap of 100 bp
* 15715 17880: contig of 2166 bp in length
* 17881 17980: gap of 100 bp
* 17981 26841: contig of 8861 bp in length
* 26842 26941: gap of 100 bp
* 26942 36150: contig of 9209 bp in length
* 36151 36250: gap of 100 bp
* 36251 62842: contig of 26592 bp in length
* 62843 62942: gap of 100 bp
* 62943 72161: contig of 9219 bp in length
* 72162 72261: gap of 100 bp
* 72262 82180: contig of 9919 bp in length
* 82181 82280: gap of 100 bp
* 82281 87039: contig of 4759 bp in length
* 87040 87139: gap of 100 bp
* 87140 104295: contig of 17156 bp in length
* 104296 104395: gap of 100 bp
* 104396 129621: contig of 25226 bp in length
* 129622 129721: gap of 100 bp
* 129722 144612: contig of 14891 bp in length
* 144613 144712: gap of 100 bp
* 144713 149573: contig of 4861 bp in length
* 149574 149673: gap of 100 bp
* 149674 153213: contig of 3540 bp in length
* 153214 153313: gap of 100 bp
* 153314 164194: contig of 10881 bp in length
* 164195 164294: gap of 100 bp
* 164295 180017: contig of 15723 bp in length.
Location/Qualifiers
1. 180017

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"

/clone="RP11-749116"
/clone_id="RP11-749116"

BASE COUNT 43238 a 43942 c 45220 g 46021 t 1596 others
ORIGIN

Alignment Scores:

Pred. No.: 2.31e-34 Length: 180017
Score: 1460.00 Matches: 463
Percent Similarity: 20.07% Conservative: 1
Best Local Similarity: 20.03% Mismatches: 7
Query Match: 45.09% Indels: 1842
DB: 2 Gaps: 4

US-10-054-935-2 (1-614) x AC068669 (1-180017)

QY 1 MetThrMetArgSerAlaValPheLysAlaAlaAlaAlaProAlaGlyAspProGlu 20
Db 66354 ATGACCATGAGATCCGGGCTTCAAGCGCGCGCCCTCGCGGCAATCTGAG 66413
QY 21 GlnArgLeuAspTyrGluArgAlaAlaAlaLeuGlyLysProGluAspGluProGlyAla 40
Db 66414 CAGCGACTGAGTACAGCGGCTGCGGCTGCGGCGCGCGAGAGAGAGCTGCGGCG 66473

QY 41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProLeuAla 60
Db 66474 GCCAAGCCCACTTCTCCCGCCGACGTAAGCTCAAGAGCCGGGCCCCGCTGACC 66533
QY 61 SerSerGlnLysLysSerProAlaProSerProAlaGlySerGlyLysGlyArgGly 80
Db 66534 TCTCTCCAGGGCGGAGACCCCGGCGCTTCCCGGCGGCTGCGGCGCAAGGGCGGGC 66593
QY 81 LeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGln 100
Db 66594 TTGTTACTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 66653
QY 101 ProLeuProCysProProProAlaThrLysGlnAlaGlyLysGlyLysGlyLysGly 120
Db 66654 CCTTGGCCCTGTCCGCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66713
QY 121 AlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProLysGlnThrGly 140
Db 66714 GCGGAGCGCGGCTGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 66773
QY 141 SerLeuValAlaAlaAlaLysGluProThrProThrAlaGlyAspLysGlyAlaAla 160
Db 66774 TTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66833
QY 161 SerProAlaAlaThrAlaSerAspProAlaGlyProProProProLeuProLeuProGlyPro 180
Db 66834 TCCCCCGCTGCCACCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 66893
QY 181 ProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlnGlyArgTyrLys 200
Db 66894 CCACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 66953
QY 201 SerMetArgLysSerProLeuGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 220
Db 66954 AGTATGAGAGAGACCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 67013
QY 221 CysLeuLysGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 240
Db 67014 TGCTCTCAACAGATCTTCTGCTCTGCTCAATGAGCTTCAACAGAGAGAGAGAGAGAG 67073
QY 241 GlnAlaLysGluLysGluLysGluLysGluLysSerGlnArgAspThr----- 256
Db 67074 CAGGCCAAGGAAAGAGAGATGAGAGACTGAAGTCAGAGAGAGAGAGAGAGAGAGAGAG 67133
QY 256 ----- 256
Db 67134 TTAAATTCATTCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67193
QY 256 ----- 256
Db 67194 GAGGTTAAGGACCCCGCGGCTTAAAGGAGTTGGCCACCGCAAGAGAGATCTCT 67253
QY 256 ----- 256
Db 67254 TAGGCGCTGGGCTCTCTGGAGCGAGGCTCAACAGACAGTGGGTTGAGAGGAAGGTTAA 67313
QY 256 ----- 256
Db 67314 AGGCAACCTCTCAGGGGAGAGGGAGAGGGTTTCAGGTGCCAAGAGGTTAATTAAATG 67373
QY 256 ----- 256
Db 67374 ACAGCCCGACAGCTGTGGGAAGACGTTTGTGGAAGAAAGACGTGATCTGTGAGGGGTTA 67433
QY 256 ----- 256
Db 67434 AAATGAAGAGGTGGGTGTAAGTGTGTGCTTATGAAGGTTAAATTAAGAGCTA 67493
QY 256 ----- 256
Db 67494 TTGCGTTAAGAAATGACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67553

QY	256	-----	256
Db	67554	TGCTGCAGACATCTTTAGATAGGAGGCGATGAGCCCTTTAAGTGGCACAAGGGGTAGAGT	67613
QY	256	-----	256
Db	67614	GATTACGTGGAATCCTATTTTAGTTTTCTTAATGATATTAGAGAAATCAGCTTTG	67673
QY	256	-----	256
Db	67674	ATTTCAGACAGCTTTTAATAGAACCTGAGTCTTAATACATGACCCAGTACAGGTCCC	67733
QY	256	-----	256
Db	67734	TTTATATAGCATGSGTGTGAACTGSAAGATTTTAGGGCATCAGGGAAAGTAAGCTTA	67793
QY	256	-----	256
Db	67794	TTAAATGTATTAAATAGACCTCCCATCCTTATTATAGAACTCTTAGAGAGTAGAAGTCC	67853
QY	256	-----	256
Db	67854	GTTTATTATTCTTTACTTCTGAGGAGCTTGGACGACACAGGCTGTGTTTACTGCAGCA	67913
QY	256	-----	256
Db	67914	GTAGTTAAGGCGATCTCGCGTTAAACAATATCTTCAGATATGAGGAGAGCTGGGCTG	67973
QY	256	-----	256
Db	67974	GGGGGTATTACTGTGCTAGCATCTTACAGCTGCCTAAATAATAAACATGGTTTCTACA	68033
QY	256	-----	256
Db	68034	TTGTTGATGTCAGGCTGCAGCAGTCTCTCTCCCTCTCTCCCTCTCTCTCTCTCTCC	68093
QY	256	-----	256
Db	68094	CTGTCTCTTCCCTTC	68153
QY	256	-----	256
Db	68154	CCTC	68213
QY	256	-----	256
Db	68214	CTATTCAATTTTGTGCTTCCAAATGAATCTCATCGGCTGTGTTTATTGGAAATG	68273
QY	256	-----	256
Db	68274	CATTCTCAAGTGGATTTGGAGCTGGGCGAGTCTCATTTAAATATCTCATGTGTGTA	68333
QY	256	-----	256
Db	68334	TTTGTGAAGTTACTCAACATGAGGAGGCGACTCTCAGACTGTACTGCAGTCTAGGGGT	68393
QY	256	-----	256
Db	68394	GTGATGCTACCGTAAATTACAACATCATATCTTGTATGGGGGAGTGGGAAGAAGG	68453
QY	256	-----	256
Db	68454	ACCGTGAAGAGAAAAAATCAAGGAATGTCCACCACAATGTGTGTTAAGGCTGTG	68513
QY	256	-----	256
Db	68514	AATCAAAACATATACAGAAGATGAATGAGAGTATGCTGATAGGGGGTTCCTGCTG	68573
QY	256	-----	256
Db	68574	ATTGTGAGGTTCTCTGCTCAGTTCTGTAGAGTAATGGCCCCCTTTGCAGGCTTCTTAGTCA	68633
QY	256	-----	256

Db	68634	GGAAAGGGGGTGGATTAGAGTAGAGTGTCAAGTCTGAACCTTGATGACACGCTGCTTGG	68693
QY	256	-----	256
Db	68694	TTGTGTATGCAATCTTAATATTGCTTAAGCTGTGTTCACTTAATTCAGTGTGTTGG	68753
QY	256	-----	256
Db	68754	CCTGTCTGAAATCTATGCAATGAATGTAAATGAAGCAGCGTTAAACATCAGTCTAAC	68813
QY	256	-----	256
Db	68814	CCAGATAAATCTTGTAAATAGAGTATAAGTAGACACAAACCTTTCTCAAACTAACT	68873
QY	256	-----	256
Db	68874	AACCAAGTTTCAATACTAATAGAGTCCATTTCTGTTAGTTTTTTAGTCTTCTCT	68933
QY	256	-----	256
Db	68934	TTTAAGTAACATAAGGAAGCTTGATTTTAATAAATATTAGATGAATAGTATAATT	68993
QY	256	-----	256
Db	68994	TTTGTGTACCTCAAAATCAAGGAATGTTATGTCAATGTTGGAAATTAAGGTA	69053
QY	256	-----	256
Db	69054	CTGATTTCTATTTCCAATGAATGTTCATGTTTAAAGTACCCCTCCACACATA	69113
QY	256	-----	256
Db	69114	CCCCTTCAATAACACTGCATCTCGTGGCAGCTATAGACTCTTAATTCCTTCA	69173
QY	256	-----	256
Db	69174	ATGTACTGTTGCACCTTGCTGTGTGTATAGACTTTGGAGTAGAAGTAGAGGATAG	69233
QY	256	-----	256
Db	69234	AATTTGGACAATCTCAGGGGCTCTTAANAAGATCATGATAGGCTGGGCGGTGGCTACGC	69293
QY	256	-----	256
Db	69294	CTGTAAATCCAGACCTTTGGAGGCTACGACAGGTGAGCACAGAGATAGAGA	69353
QY	256	-----	256
Db	69354	CCATCTGGCCAAACATGTTGAACCTGTCTCTACTATAAATATCAAAAATTTACCGGTGT	69413
QY	256	-----	256
Db	69414	GGTGGGGGCGCTGAATAATCCAGCTACTCGGAGGCTGAGCGAGGAGCGGAGGTTGC	69473
QY	256	-----	256
Db	69474	AGTAGCCATTGCATCTCAGCCTTGCGCACAAAGCAGCAAACTCCGTCTCAAAAACAAAAC	69533
QY	256	-----	256
Db	69534	AAAAACAACAAACAAACATGTATACAATTAAATTCTGGAAACTCCGAAACCAACTTGA	69593
QY	256	-----	256
Db	69594	GCCACACAGGTAGCAGTAGCAGAGACCATGCATTAAAGGATGTTACGGCCCTTCCCA	69653
QY	256	-----	256
Db	69654	GTGGAGAGGATCTCCCTACAGAAAGATGAACACTCAAAATTACCCCATGCTCTTCCCATCT	69713
QY	256	-----	256

Db 69714 AGGTTAGCCCTTAATAATTTAACAGTATACCTTCTCTCTATTTGTTAAGGAAACT 69773
QY 256 ----- 256
Db 69774 GTTCAGATTATGCTCTCTATTATAAAGTTAGTTGATGTGCTCCATGAAAGAGCA 69833
QY 256 ----- 256
Db 69834 GAAGACTATACTGATTCCTTAATGGGCTAAGTATCTGTGCTTATTTTATGTGTT 69893
QY 257 ----- LeuLeuLlaarqll egluaqmetluarg 266
Db 69894 AAGTCTCATTTTCTCTCTCTCTTTTAACTCTCTCTCTGATGGAACGATGGAAGG 69953
QY 267 ArgmetGlnLeuValLysLysAspAsnGlnLysGlnurghLysLysLeuPheGlnGlyTyr 286
Db 69954 CGGATGACGCTGGTAAGAAAGATACGAGAAAGAAAGCAAGCTGTTCACAGGCTAT 70013
QY 287 GluThrGlnLuarargLuglLurghLulLeuSerGlnLysLleLysLeuGluCySglPro 306
Db 70014 GAAACTGAAGAGAGAGAGAAACAGACTATCTGAGAAATTAACCTGGAGTGCACGCC 70073
QY 307 GluLeuSerGlnurhSerGlnurhLeuProProlysProPheSerCysGlyArgSerGly 326
Db 70074 GAGCTTCCGAGACATCCAGACTCTGCTCCCAAGCCCTCTCATGTGGCGGAGTGA 70133
QY 327 LysGlyHisLys ----- 330
Db 70134 AAGGACATAA - AAGGTGCTGATAACTTTTGATGAGAGACCCTGTTACAGATGAT 70192
QY 330 ----- 330
Db 70193 TGAGAGTAGGAGATTGGGATTATGANTGTTTATAGCAATATACTTGTTAGAA 70252
QY 330 ----- 330
Db 70253 GTCTTATGCGAGAGCGCGGTGTGTGCTCACACTGTAATCCAGACTTTGGAG 70312
QY 330 ----- 330
Db 70313 GCCAAGCGGGGTGATCATGAGTCAGAGATAGAGACCATCTGGCCACATTTTGAA 70372
QY 330 ----- 330
Db 70373 CCCCCTCTCTATAAATATATAAATTTAGCCGGCGGTGATGGCGGCGCTGTAGTCC 70432
QY 330 ----- 330
Db 70433 CAGCTACTCGGAGGCTGAGGAGAGAAATCATCTTGAACCGGAGGCGAGTTGAGT 70492
QY 330 ----- 330
Db 70493 GAGCGGGGATTGCTCCACTGACTCCAGCTCGGACAGAGAGGAGACTCCGTCAAAAA 70552
QY 330 ----- 330
Db 70553 AAAAAAAGAAGTCTTCTATGCTGAGAAACTATCTGAAACATTTCTTATGACACTGTA 70612
QY 330 ----- 330
Db 70613 AAAGCTGTTAGAGATTTTGTGTCACAGTAATATATTGAATGACAGTTACATCAT 70672
QY 330 ----- 330
Db 70673 AAGGAATGGCAATTAAGTATGACAGCTGAGTTCTTAAACTATAGCAGTTCACTTCAA 70732
QY 330 ----- 330
Db 70733 GTTAGGAAGTGTAAATTCCTTCAATGTGTGCTGATTTGTAGACATAGAAAAGTAAT 70792
QY 330 ----- 330
Db 70793 ACAGGCAAGCGGTGTGCTCATGCTGTAACTCCCTTTGAGAGGGCAAGGTGGGAAGA 70852

QY 330 ----- 330
Db 70853 GTATGAAAGTATGAAAGTGGGAAGAGAGGATCGCTTGCCCGCAGAGATTGAGACCA 70912
QY 330 ----- 330
Db 70913 GCCTGGCAACATFAGTAGACATCTTTTACAAAAAATCCAAAAATTAGCGGGCGCTGG 70972
QY 330 ----- 330
Db 70973 AGTATGCACTGTAGTCCAGACTACTAGAGGTTGAGGCTACAGTGAAGCTGTACTGC 71032
QY 330 ----- 330
Db 71033 ACCACTGACTCCAGCCTGGGTGACGAGTATACCTGTCTCAATTGAAAAAATAA 71092
QY 330 ----- 330
Db 71093 AAAAAGTAAATGAACGTCTCTTGAGACTCAAAAAAATTAACCCAACTGTAAAGTAG 71152
QY 330 ----- 330
Db 71153 GCATACATTTTGAACACTCAAAATTCAAATGCAAGTAAATACGTAAATGACACACCTG 71212
QY 330 ----- 330
Db 71213 ATATAACTAAGCTTTTGTGTTGTGATAGGCTCTCTGTGACCCAGGCTGGG 71272
QY 330 ----- 330
Db 71273 GTGCAAGTGTGCCAATATGCTCATTCACTTCACTTCCACTGTGAGCTCAAGTATCTC 71332
QY 330 ----- 330
Db 71333 CTACTCAGCTCCCATGTAGCTAGAGACCATAGGCACTCACACATGTTGCTAATTTT 71392
QY 330 ----- 330
Db 71393 TTAGTTTTGTAGAGACAGGCTCTCACTTTGTTTCCAGGCTGATCTCAAACTCTTGCGC 71452
QY 330 ----- 330
Db 71453 TCAGTGATCTCCCGAAGTGTGGATTTACAGACATTAACATAAGTTTGTAAAGTCA 71512
QY 330 ----- 330
Db 71513 GGGATATTAAAGTTATTTGTGGTTAGAAATTTGAAACTTTAAGTAGAAATCACTTTGA 71572
QY 330 ----- 330
Db 71573 AAAAAGTATAGCCGGGTGCGGTGCTCACGCTGTAAATCCAGCACTTTGGGAGCGG 71632
QY 330 ----- 330
Db 71633 AGGCGGTGATTCAGAGGTGAGAGATCGAGACATCTGTGTAACACGTTGAACCT 71692
QY 330 ----- 330
Db 71693 GTCTTACTAAAAATAGAAAAATTAAGCCGGGTGTGTCATACACTGTATGCCAGCTA 71752
QY 330 ----- 330
Db 71753 CTTGGAGGCTGAGGAGAGAAATGCTTGAACCCAGAGGTGAGGTTGAGTAAGCAG 71812
QY 330 ----- 330
Db 71813 AGATCGACACATGCACTCCACTCTCGGGTGACAGAGTGAAGCTCCAAAAAAGTATTA 71872
QY 330 ----- 330
Db 71873 TTTAGTTGTAAAGAGAAAGGAGATGACTTTCTTGGATGTGTGGGGAAAGAGGA 71932

QY 330 ----- 330
 Db 71933 AGAAGAAAGAGCTAGAGATGCCAGACAGATGAAGAAAAGAAATTAAGTTACAT 71992
 QY 330 ----- 330
 Db 71993 TAATGTGTACAGGGTTGGGGGAAAGAAAGAGACTGATCTGTGATTACAGAGTAG 72052
 QY 330 ----- 330
 Db 72053 GTCCACCTATGTGGGTTGTAGTGTACATATGTTGGCTAAAGACCTTGATTTCTT 72112
 QY 330 ----- 330
 Db 72113 TTTTCTTTTCTGTGAGACAGAGCTCAGTCTGTGCCCCANNNNNNNNNNN 72172
 QY 330 ----- 330
 Db 72173 NNN 72232
 QY 330 ----- 330
 Db 72233 NNN 72292
 QY 330 ----- 330
 Db 72293 TAATCTTAACATGAAATTCCTTGTTCAACATTTACTGCGCAGTGCACTGCTCATGC 72352
 QY 330 ----- 330
 Db 72353 CGGTATCCAGACACTTTGGGAGGCGTAGTGGGTGATCCTGAGTCAAGTATTTGA 72412
 QY 330 ----- 330
 Db 72413 GACCAGTGTGACCAATATGTGTGAACCCCTTTTATTAATAATACAAAAAATTAC 72472
 QY 330 ----- 330
 Db 72473 AGGCATGTGTGTATGCTGTAATCCAGCTACTCAGAGGCTGAGATGAGAAATTC 72532
 QY 330 ----- 330
 Db 72533 TTGACCCGGAGGAGAGATGTCAGTGCAGCCAGATCATGCTCCACTCCAGCTTGG 72592
 QY 330 ----- 330
 Db 72593 GTGACAGAGACACTGTGTCTCAAAAACAACAAAACAAAATTTACTATGG 72652
 QY 330 ----- 330
 Db 72653 GCCTATTTTATATCTGCTAAGCCCTGGCTGTTCCCTGCTCATGCTAATGAGCACC 72712
 QY 330 ----- 330
 Db 72713 AGACATTTAGCTACCTGATCTGTCATTTTGGGGCAAAAAGAACCGCTTATTTTG 72772
 QY 330 ----- 330
 Db 72773 CTCACAGATGTTCAAATTAAGATTTTATCAAGTGTCAATTAATTTATATGTTTC 72832
 QY 331 ----- 331
 Db 72833 TCCCTATCTAATAGGAATCCCATTTGGAGTACAGAAAAGAAAGACTCTCTGTTAAAG 72892
 QY 347 ----- 347
 Db 72893 CTGGCTCTCAATTTTCAAAAGTCAAAACAAACTCTTAAGCACTCTCTATTAAGAG 72952
 QY 367 ----- 367
 Db 72953 GAACCCGTGTCTTATCTGAACGTTTGTAAACGTGAATTTGAGAGCAAGAAC 73012
 QY 387 ----- 387
 Db 387 ProgluylsProalrSerSerValAspThrProalrGlueSerThrProgluylsGly 406

Db 73013 CCAGAAAAGCCCGGCTTCAGTGGACACCCACCAAGACTCTCCACTCCCAAGAGGA 73072
 QY 407 ----- 407
 Db 73073 CCCAGACCCCATCCCAAGGAGAAAGCCCTCTCAAGTGAAGATTAAGATTTGGCGTACCTT 73132
 QY 427 ----- 427
 Db 73133 TCCACCAAGAAATGATTTGTCTGTTGGCACACCTCCCATCCACCGTTACCATTA 73192
 QY 447 ----- 447
 Db 73193 CGGGAATCCCTCTCAAGAGAGGAGAGACTGTAGCA-----ACTAAG 73234
 QY 467 ----- 467
 Db 73235 GCATAGAGACACTTGTCTT---ATACCTTAATGG 73267
 RESULT 3
 AC119462
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-28N16, *** SEQUENCING IN PROGRESS
 ***79 unordered pieces.
 AC119462
 VERSION
 KEYWORDS
 HTG: HTGS_PHASE1.
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 193189)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaratunga,H.C., Are,J.R., Ayala,M., Banks,T.,
 Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burckell,C., Burrell,K.L., Byrd,N.C.,
 Carmon,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,K.J.,
 Doultwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huily,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kravovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louisedge,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newson,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogund,H., Okwoum,G.,
 Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojuben,I., Rolfe,M., Ruiz,S., Severy,G.,
 Scherer,S., Scott,G., Shen,R., Shoshari,N., Stinson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Washington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
 Weinstock,G., and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193189)
AUTHORS Morley,K.C.
JOURNAL Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 193189)
Morley,K.C.
REFERENCE Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20387269.
COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GURK
Center clone name: CH230-28N16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 126440 bases at least Q40
Consensus quality: 140831 bases at least Q30
Consensus quality: 148852 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1554: contig of 1554 bp in length
* 1555 1654: gap of unknown length
* 1655 2715: contig of 1061 bp in length
* 2716 2815: gap of unknown length
* 2816 3899: contig of 1084 bp in length
* 3900 3999: gap of unknown length
* 4000 5367: contig of 1368 bp in length
* 5368 5467: gap of unknown length
* 5468 6690: contig of 1223 bp in length
* 6691 6790: gap of unknown length
* 6791 8012: contig of 1222 bp in length
* 8013 8112: gap of unknown length
* 8113 9566: contig of 1454 bp in length
* 9567 9666: gap of unknown length
* 9667 10914: contig of 1248 bp in length
* 10915 11014: gap of unknown length
* 11015 12395: contig of 1381 bp in length
* 12396 12495: gap of unknown length
* 12496 14050: contig of 1555 bp in length
* 14051 14150: gap of unknown length
* 14151 15648: contig of 1498 bp in length
* 15649 15748: gap of unknown length
* 15749 16994: contig of 1246 bp in length
* 16995 17094: gap of unknown length
* 17095 18274: contig of 1180 bp in length
* 18275 18375: gap of unknown length
* 18376 20426: contig of 2052 bp in length
* 20427 20526: gap of unknown length
* 20527 22026: contig of 1500 bp in length
* 22027 22126: gap of unknown length
* 22127 23257: contig of 1131 bp in length
* 23258 23357: gap of unknown length

23358 24652: contig of 1295 bp in length
* 24653 24753: gap of unknown length
* 24753 26187: contig of 1435 bp in length
* 26188 26287: gap of unknown length
* 26288 27567: contig of 1280 bp in length
* 27568 27667: gap of unknown length
* 27668 28980: contig of 1313 bp in length
* 28981 29080: gap of unknown length
* 29081 30136: contig of 1056 bp in length
* 30137 30236: gap of unknown length
* 30237 31633: contig of 1396 bp in length
* 31633 31733: gap of unknown length
* 31733 33270: contig of 1538 bp in length
* 33271 33371: gap of unknown length
* 33371 35465: contig of 2095 bp in length
* 35466 35565: gap of unknown length
* 35566 37201: contig of 1636 bp in length
* 37202 37301: gap of unknown length
* 37302 38947: contig of 1646 bp in length
* 38948 39047: gap of unknown length
* 39048 40710: contig of 1663 bp in length
* 40711 40810: gap of unknown length
* 40811 42463: contig of 1653 bp in length
* 42464 42564: gap of unknown length
* 42564 43912: contig of 1349 bp in length
* 43913 44012: gap of unknown length
* 44013 45591: contig of 1579 bp in length
* 45592 45691: gap of unknown length
* 45692 47620: contig of 1929 bp in length
* 47621 47720: gap of unknown length
* 47721 49829: contig of 2109 bp in length
* 49830 49929: gap of unknown length
* 49930 51465: contig of 1536 bp in length
* 51466 51565: gap of unknown length
* 51566 52865: contig of 1300 bp in length
* 52866 52965: gap of unknown length
* 52966 54343: contig of 1378 bp in length
* 54344 54443: gap of unknown length
* 54444 55895: contig of 1452 bp in length
* 55896 55995: gap of unknown length
* 55996 57574: contig of 1579 bp in length
* 57575 57674: gap of unknown length
* 57675 59301: contig of 1627 bp in length
* 59302 59401: gap of unknown length
* 59402 61393: contig of 1992 bp in length
* 61394 61493: gap of unknown length
* 61494 63075: contig of 1582 bp in length
* 63076 63175: gap of unknown length
* 63176 64817: contig of 1642 bp in length
* 64818 64917: gap of unknown length
* 64918 67402: contig of 2485 bp in length
* 67403 67502: gap of unknown length
* 67503 69777: contig of 2275 bp in length
* 69778 69877: gap of unknown length
* 69878 71982: contig of 2105 bp in length
* 71983 72082: gap of unknown length
* 72083 74289: contig of 2207 bp in length
* 74290 74389: gap of unknown length
* 74390 76937: contig of 2548 bp in length
* 76938 77037: gap of unknown length
* 77038 79003: contig of 1966 bp in length
* 79004 79103: gap of unknown length
* 79104 81519: contig of 2416 bp in length
* 81520 81619: gap of unknown length
* 81620 84864: contig of 3245 bp in length
* 84865 84964: gap of unknown length
* 84965 87539: contig of 2575 bp in length
* 87540 87639: gap of unknown length
* 87640 89437: contig of 1798 bp in length
* 89438 89537: gap of unknown length
* 89538 92108: contig of 2571 bp in length
* 92109 92208: gap of unknown length
* 92209 94902: contig of 2694 bp in length

```

*      94903      95002: gap of unknown length
*      95003      98082: contig of 3080 bp in length

Alignment Scores:
Pred. No.:      6,62e-30      Length:      193189
Score:          1311.00      Matches:      254
Percent Similarity: 94.91%      Conservative: 7
Best Local Similarity: 92.36%      Mismatches: 10
Query Match:    40.49%      Indels:      4
DB:             2      Gaps:      2

US-10-054-935-2 (1-614) x AC119462 (1-193189)

QY      1 MetThMetArSerAlaValPheLysAlaAlaAlaProAlaGlyAsnProGlu 20
      160970 ATGACCATGATGATCCGACAGTTCACAGCGCGCGCCCTCCGCGGCAACCCCGAG 161029
      |||
QY      21 GlnArgLeuAspTyrGluArgAlaAlaLeuGlyGlyProGluAspGluProGlyAla 40
      161030 CACGACCTGACATACAGCGGCTGCGGCTGCGGCGCCAGAGACGATCCGCGGCGG 161089
      |||
QY      41 AlaGluAlaHisPheLeuProArgHisArgLysLeuLysGluProGlyProProLeuAla 60
      161090 GCGGAACCCACATCTCTCCCGCGCATCTGATGACTCAAGAGCGCGCGCCCGCTGCGCC 161149
      |||
QY      61 SerSerGlnGlyLysSerProAlaProSerProAlaGlyCys---GlyGlyLysGlyArg 79
      161150 TCTACCCAGGCGCGGAGCCCGCGCTCTCCAGCGGCTGCGGCGCGGCGCAAGGCGCGG 161209
      |||
QY      80 GlyLeuLeuLeuProAlaGlyAlaAlaProGlyGlnGlnGluLeuSerTyrGlyLysSer 99
      161210 GGTTCGTACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161269
      |||
QY      100 ValProLeuProCysProProProAlaThrLysGlnAlaGlyLysGlyGluProAla 119
      161270 GTCCCTTGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161329
      |||
QY      120 AlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuProLysGlnThr 139
      161330 GCAGCGCGCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161389
      |||
QY      140 GlySerLeuValAlaAlaAlaLysGluProThrProThrAlaGlyAspLysGlyAla 159
      161390 GGGTCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161449
      |||
QY      160 AlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuProLeuProGly 179
      161450 GCTGCCCGACGCTGCCAGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161509
      |||
QY      180 ProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGluGlyArgTyr 199
      161510 CCGCCACCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 161569
      |||
QY      200 LysSerMetArgLysSerProLeuGlyGlyGlyGlySerLysAlaSerSerGlnAla 219
      161570 AAGAGATATAGAGAGAGCCCTCTCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 161629
      |||
QY      220 AlaCysLeuLysGlnLeuLeuLeuGlnLeuAspLeuLysGluGlnGlnGlnGln 239
      161630 GCGTGGCTCAAAAGATCTCTGCTGCTGCAATTGCGACTCATCGAACAGCAGCAGCAGC 161689
      |||
QY      240 LeuGlnAlaLysGluLysGluLysGluLysSerGluArgAspThrLeu----- 257
      161690 TTGCAAGCCCAAGAGAGAGAGATCGAGAGCTGAACTCGAGAGATACGATACGAGGAG 161749
      |||
QY      258 ---LeuAlaArgLysGluArgMetGluArgArgMetGlnLeuVal 271
      161750 GGGTAAATCTGCAATAGGATGGGAGGAGCGACTCTGCGCATGCTC 161794
      |||
RESULT 4
AX401049 615 bp DNA linear PAT 06-JUN-2002
LOCUS Sequence 725 from Patent WO0210453.
DEFINITION

```

```

ACCESSION AX401049
VERSION AX401049.1 GI:21337229
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE Mendrick,D., Porter,M.W., Johnson,K.R., Castle,A.L. and
AUTHORS Elashoff,M.R.
TITLE Molecular toxicology modeling
JOURNAL Patent: WO 0210453-A 725 07-FEB-2002;
FEATURES Gene Logic, Inc. (US)
source location/Qualifiers
1.615
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/note="EMBL/Genbank Accession No. A1102093"
BASE COUNT 84 a 235 c 224 g 71 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      1.99e-24      Length:      615
Score:          1055.50      Matches:      197
Percent Similarity: 97.55%      Conservative: 2
Best Local Similarity: 96.57%      Mismatches: 4
Query Match:    32.60%      Indels:      1
DB:             6      Gaps:      1

US-10-054-935-2 (1-614) x AX401049 (1-615)

QY      13 AlaProAlaGlyLysLysAsnProGluGlnArgLeuAspTyrGluArgAlaAlaLeuGly 32
      3 GCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62
      |||
QY      33 GlyProGluAspGluProGlyAlaAlaGluAlaHisPheLeuProArgHisArgLysLeu 52
      63 GCGCGCGAGAGACAGTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
      |||
QY      53 LysGluProGlyProProLeuAlaSerSerGlnGlySerProAlaProSerProAla 72
      123 AAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
      |||
QY      73 GlyCys---GlyLysGlyArgGlyLeuLeuProAlaGlyAlaAlaProGlyGln 91
      183 GCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 242
      |||
QY      92 GlnGluGluSerTyrProLysSerValProLeuProCysProProProAlaThrLysGln 111
      243 CAGGAAGAGAGCGTGGGGCGGTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
      |||
QY      112 AlaGlyLeuGlyLysGluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyr 131
      303 GCCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
      |||
QY      132 GlnAlaValLeuProLysGlnThrGlySerLeuValAlaAlaLysGluProThrPro 151
      363 CAGCGGCTGTGCGCATTCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 422
      |||
QY      152 TrpAlaGlyAspLysGlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGly 171
      423 TGGGCTGGGAGACAGAGGTGGGGGCTGCCCGCGCTGCGCACCGCTCGGAGCCGCGGGA 482
      |||
QY      172 ProProProLeuProLeuProGlyProProProProAlaProThrAlaThrAlaGlyThr 191
      483 CCGCCACCTACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542
      |||
QY      192 LeuAlaAlaSerGluLysArgTyrLysSerMetArgLysSerProLeuGlyLysGly 211
      543 CTGGCGCGCAGAGGAGGAGAGATGAGAACTATTAAGAAAGAGCGCTCTCGGGGTGGCGCG 602
      |||
QY      212 GlySerGlyAla 215

```

DB 603 NGCTCGGAGCC 614

|||||||

RESULT 5 AC131356/c 174346 bp DNA linear HTG 21-AUG-2002

LOCUS Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS

DEFINITION *** 55 unordered pieces.

ACCESSION AC131356.1 GI:22380622

VERSION AC131356

KEYWORDS HTG: HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1. (bases 1 to 174346)

AUTHORS Muzny,D,Marle,M, Metzger,M, Lee,A, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angiano,D, Anyalbechli,Y, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,Y, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedetich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Flinley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Gara,M, Gengeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guetara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hernandez,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpaty,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowls,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenschewa,L, Louiseged,H, Lozano,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Marhiley,S, McLeod,M, McNeill,T, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwackeleme,O, Okunodu,G, Olarunpusagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Poldexter,A, Popovic,D, Primus,E, Pu,L, Puazo,M, Qutro,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,I, Shatsman,S, Shen,H, Shetty,J, Shvartbeyan,A, Sisson,L, Sitter,C,D, Smajd,D, Sneed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Syatek,A, Tabor,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trefos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,S, Wang,S, Warren,R, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wlaczek,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (21-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web Site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: 600J

Center clone name: CH230-278A19

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 109854 bases at least Q40

Consensus quality: 11811 bases at least Q30

Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 55 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1

1547: contig of 1547 bp in length

1548

1647: gap of unknown length

1648

3064: contig of 1417 bp in length

3065

1648: gap of unknown length

3165

4885: contig of 1721 bp in length

4886

4985: gap of unknown length

4986

6000: contig of 1015 bp in length

6001

6000: gap of unknown length

6101

7171: contig of 1071 bp in length

7172

7271: gap of unknown length

7272

8794: contig of 1523 bp in length

8795

8894: gap of unknown length

8895

10368: contig of 1474 bp in length

10369

10468: gap of unknown length

10469

1306: contig of 1438 bp in length

11907

12006: gap of unknown length

12007

13031: contig of 1025 bp in length

13032

13131: gap of unknown length

13132

14575: contig of 1444 bp in length

14576

14675: gap of unknown length

14676

15699: contig of 1024 bp in length

15700

16951: gap of unknown length

16951

16951: contig of 1152 bp in length

16952

17051: gap of unknown length

17052

16695: contig of 1644 bp in length

16696

18795: gap of unknown length

18796

20470: contig of 1675 bp in length

20471

20570: gap of unknown length

20571

22418: contig of 1448 bp in length

22419

22518: gap of unknown length

22519

24364: contig of 1846 bp in length

24365

24464: gap of unknown length

24465

25586: contig of 1122 bp in length

25587

25686: gap of unknown length

25687

27859: contig of 2173 bp in length

27860

27959: gap of unknown length

27960

29649: contig of 1690 bp in length

29650

29749: gap of unknown length

29750

31331: contig of 1362 bp in length

31332

31431: gap of unknown length

31432

33706: contig of 2175 bp in length

33707

35179: contig of 1473 bp in length

35180

35279: gap of unknown length

35280

37193: contig of 1914 bp in length

37194

37293: gap of unknown length

37294

38382: contig of 1089 bp in length

Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunolu, G.,
 Oringye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Qules, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
 Soderren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swalek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, D., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 193189)
 Worley, K. C.
 Direct Submission
 Submitted (27-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193189)
 Worley, K. C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20387269.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GURK
 Center clone name: CH230-28N16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 126440 bases at least Q40
 Consensus quality: 140831 bases at least Q30
 Consensus quality: 148852 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 79 contigs. The true order of the record is
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1554: contig of 1554 bp in length
 * 1555 1654: gap of unknown length
 * 1655 2715: contig of 1061 bp in length
 * 2716 2815: gap of unknown length
 * 2816 3899: contig of 1084 bp in length
 * 3900 3999: gap of unknown length
 * 4000 5367: contig of 1368 bp in length
 * 5368 5467: gap of unknown length
 * 5468 6690: contig of 1223 bp in length
 * 6691 6790: gap of unknown length
 * 6791 8012: contig of 1222 bp in length
 * 8013 8112: gap of unknown length
 * 8113 9566: contig of 1454 bp in length
 * 9567 9666: gap of unknown length.
 * 9667 9667: contig of 1248 bp in length
 * 10915 11014: gap of unknown length
 * 11015 12395: contig of 1381 bp in length
 * 12396 12495: gap of unknown length
 * 12496 14050: contig of 1555 bp in length
 * 14051 14500: gap of unknown length
 * 14501 15548: contig of 1498 bp in length
 * 15549 15749: gap of unknown length
 * 15749 16994: contig of 1246 bp in length
 * 16995 17095: gap of unknown length
 * 17095 18275: contig of 1180 bp in length
 * 18275 18375: gap of unknown length
 * 18375 20427: contig of 2052 bp in length
 * 20427 20527: gap of unknown length
 * 20527 22026: contig of 1500 bp in length
 * 22026 22126: gap of unknown length
 * 22126 23257: contig of 1131 bp in length
 * 23257 23357: gap of unknown length
 * 23357 24652: contig of 1295 bp in length
 * 24652 24752: gap of unknown length
 * 24752 26187: contig of 1435 bp in length
 * 26187 26287: gap of unknown length
 * 26287 27567: contig of 1280 bp in length
 * 27567 27667: gap of unknown length
 * 27667 28980: contig of 1313 bp in length
 * 28980 29080: gap of unknown length
 * 29081 30136: contig of 1056 bp in length
 * 30137 30236: gap of unknown length
 * 30237 31632: contig of 1396 bp in length
 * 31633 31732: gap of unknown length
 * 31733 33270: contig of 1538 bp in length
 * 33271 33370: gap of unknown length
 * 33371 35465: contig of 2095 bp in length
 * 35466 35565: gap of unknown length
 * 35566 37201: contig of 1636 bp in length
 * 37202 37301: gap of unknown length
 * 37302 38947: contig of 1646 bp in length
 * 38948 39048: gap of unknown length
 * 39049 40710: contig of 1663 bp in length
 * 40711 40810: gap of unknown length
 * 40811 42463: contig of 1653 bp in length
 * 42464 42563: gap of unknown length
 * 42564 43912: contig of 1349 bp in length
 * 43913 44012: gap of unknown length
 * 44013 45591: contig of 1579 bp in length
 * 45592 45691: gap of unknown length
 * 45692 47620: contig of 1929 bp in length
 * 47621 47720: gap of unknown length
 * 47721 49829: contig of 2109 bp in length
 * 49830 49929: gap of unknown length
 * 49930 51465: contig of 1536 bp in length
 * 51466 51565: gap of unknown length
 * 51566 52865: contig of 1300 bp in length
 * 52866 52965: gap of unknown length
 * 52966 54343: contig of 1378 bp in length
 * 54344 54443: gap of unknown length
 * 54444 55895: contig of 1452 bp in length
 * 55896 55995: gap of unknown length
 * 55996 57574: contig of 1579 bp in length
 * 57575 57674: gap of unknown length
 * 57675 59301: contig of 1627 bp in length
 * 59302 59401: gap of unknown length
 * 59402 61393: contig of 1992 bp in length
 * 61394 61493: gap of unknown length
 * 61494 63075: contig of 1582 bp in length
 * 63076 63175: gap of unknown length
 * 63176 64817: contig of 1642 bp in length
 * 64818 64917: gap of unknown length
 * 64918 67402: contig of 2485 bp in length
 * 67403 67502: gap of unknown length
 * 67503 69777: contig of 2275 bp in length
 * 69778 69877: gap of unknown length
 * 69878 71982: contig of 2105 bp in length

COMMENT

REFERENCE
AUTHORS
TITLE
JOURNALTITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

* 71983 72082: gap of unknown length
* 72083 74289: contig of 2207 bp in length
* 74289 74389: gap of unknown length
* 74390 76937: contig of 2548 bp in length
* 76938 77037: gap of unknown length
* 77038 79003: contig of 1966 bp in length
* 79004 79103: gap of unknown length
* 79104 81519: contig of 2416 bp in length
* 81520 81620: gap of unknown length
* 81620 84864: contig of 3245 bp in length
* 84865 84965: gap of unknown length
* 84965 87540: contig of 2575 bp in length
* 87540 89437: gap of unknown length
* 89437 89538: contig of 1798 bp in length
* 89538 92108: contig of 2571 bp in length
* 92108 92209: gap of unknown length
* 92209 94503: contig of 2694 bp in length
* 94503 95003: gap of unknown length
* 95003 98082: contig of 3080 bp in length

```

Alignment Scores:

```

Pred. No.: 7.09e-18 Length: 193189
Score: 906.50 Matches: 221
Percent Similarity: 43.60% Conservative: 14
Best Local Similarity: 41.00% Mismatches: 55
Query Match: 28.00% Indels: 249
DB: 2 Gaps: 8

```

US-10-054-935-2 (1-614) x AC119462 (1-193189)

```

QY 149 ProthProtpAlaGlyAspLysGlyGlyAlaAla-SerProAlaAlaThrLaserAs 168
DB 186401 CCGGACAGGCGGAGATCATCTTCAGAGAGATGATCTP----- 186362
QY 168 ProAlaGlyProProleuProleuProGlyProProProleuAlaPro-----Th 186
DB 186361 -CGGAGTGGTCCATCCCTATTACCATCCAGGTTAACTTCATATTCATGCTTTT 186303
QY 186 rAlaThrAlaGlyThrLeuAlaAlaLaserGluGlyArgTrpLysSerMetArgLysSerPr 206
DB 186302 GCCTTACGCGGTTTAAAGGAACTATTCAGTGTGCTGCCATTCAGAGAAAGTTG 186243
QY 206 oleuGly-GlyGlyGlyGlySerGlyAlaSerSerGlnAlaAlaCysLeuLysGln---- 224
DB 186242 AATTGGTAGGG-----TCAACAGCAGCATGTAGAACGCCCAATT 186204
QY 225 -----IleLeuLeuGlnLeuAlaPleuIleGluG 235
DB 186203 CCTACATGGGGTTAACTGCTGCTAATTTCTATGATTAAGCTGCACTGTGTGACTT 186144
QY 235 InGlnGlnGlnLeuGlnAlaLysGluLysGluIleGluLysSerGluArgA 255
DB 186143 CC-----T 186141
QY 255 spThrLeuLeuAlaAaGllleGluArgMetGluArgArgMetGlnLeuValLysLysAspA 275
DB 186140 TTATGCTCTGCTCGATTCGATTGAACGTATGAAAGCGGATGCGATGCTGCTAAAGGATA 186081
QY 275 snGluGlyGluArgHsLysLeuPheGlnGlyArgGluThrGluGluArgGluGluThrG 295
DB 186080 ACAGAGAGGAAAGGCAACAGCTTTTCAGGGCTATGAACTGTAAGAGAGAGAGAGAAACGG 186021
QY 295 IuLeuSerGluLysIleLysLeuGluCysGlnProGluLeuSerGluThrSerGlnThrL 315
DB 186020 AGTTGCTGAGAAATTAACCTGAGAGCCAGCGGAGCTCTCGAGACATCCAGAGCTC 185961
QY 315 euProLysPProPheSerCysGlyArgSerGlyLysGlyHsLys----- 330
DB 185960 TGCTTTCTAAGCCTTTCTCTCATGTGCGCGAGTGGCAAGGAGACAAAGGTGTGTGTG 185901
QY 330 ----- 330

```

```

DB 185900 ACTTATTTCTATGAACTGAAAAAGATGAGAACTGCTAGTGTGATGAAATATGCGCCTT 185841
QY 330 ----- 330
DB 185840 GATCTGTGATTTGGCTTTAGACACACCATAGCCTGAATTTGAGAAAGTCTGTGTACTGGA 185781
QY 330 ----- 330
DB 185780 GCATGTGGCTTCCCTGTATATCAGACACTAGTTCAGAGGCGAGCCTGGCAATAGAGAA 185721
QY 330 ----- 330
DB 185720 CCTGTTTTTACTGTGCTTGAGACATTTTTTTTGTACAGTCTTAAGTACAGAAATGCA 185661
QY 330 ----- 330
DB 185660 GAGTATATTTTAAAGAAATGGAATAGGGGTTGGGGATTTTACTCTAGTGTAGACGCT 185601
QY 330 ----- 330
DB 185600 TGCTTAGCAAGCTCAAGGCCCTGGTTCGGTCCCGACGTCCGAAAAAAGAACCAAAAAA 185541
QY 330 ----- 330
DB 185540 AAAAAAAGAAATAGTCTTATAGCATAGACGCTTGTCAATGACCAAAAGGCTTTCTGTA 185481
QY 330 ----- 330
DB 185480 TACTTTAAGCATGACACTTACTGTCTCATGCTTCCAAATGAAACATGATATTTAGATA 185421
QY 330 ----- 330
DB 185420 CCTGGCTACCTCTTGTGCCCTCAATATATTCAAATCAAAACTTATTTGAATTAATTA 185361
QY 331 -----ArgLysSerProPheGlySerThrGluArgL 341
DB 185360 TTTTATTAACTTTTCTTTCGTACCTTAATATGAGAAACCCATTTGGAATATACAGAAAGAA 185301
QY 341 ysthrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrProLysH 361
DB 185300 AGACTCCTGTTTAAAGCTGGGCTCTGAAATTTCAAAATTCAAAACCAAAACCTCTTAAGC 185241
QY 361 lsserProIleLysGluGluProCysGlySerLeuSerGluThrValLysLysArgGluL 381
DB 185240 ACTCTCCCAATTAAGGAGGAGCCCTGTGTTCCATATCAGAAACTGTTTGTAAACGTGAAT 185181
QY 381 euArgSerGlnGluThrProGluLysProArgSerSerValAspThrProProArgLys 401
DB 185180 TGAAGAGCCAGAAACCCAGAAAGCCCGGCTTCATGATGATACCCACCAAGACTCT 185121
QY 401 erThrProGluLysGlyProSerThrHisProLysGluLysAlaPheSerSerGluIleG 421
DB 185120 CGACTCCCAAAAGGAGACCCAGCACCCACCCCAAGAGAGCAAGCTCTCAAGTAGATG 185061
QY 421 lnaSpluProtyrLeuSerThrThrGluMetLysLeuLysArgGlyPheGlnProProP 441
DB 185060 AAAGATTGGCGTACCTTTCCACACAGAAATGATTTGTCTGTGGCACACCTCCGCC 185001
QY 441 roserProleuProleuArgGluSerSerProLysLysGluGluThrValAlaArgCysL 461
DB 185000 CATCACCGTTTACCATTAACGGGAATCTCTCCAAAGAGAGAGAGACTGTAGCA----- 184948
QY 461 euMetProSerSerValAlaGlyGluThrSerValLeuAlaValProSerTrp 478
DB 184947 -----AGTAAGCATATACGAACACTTCTCTT---ATACCTAGTGG 184909

```

```

RESULT 7
LOCUS HSM800235 1433 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFp586B1922 (from clone DKFp586B1922).
ACCESSION AL049450
VERSION AL049450.1 GI:4500236
KEYWORDS

```



```
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14249
Center clone name: 54.L_14
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 200414 bases at least Q40
Consensus quality: 203403 bases at least Q30
Consensus quality: 204607 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 205214; sum-of-contigs
Quality coverage: 7.0 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 9177: contig of 9177 bp in length
9178 9277: gap of 100 bp
9278 9931: contig of 654 bp in length
9932 10031: gap of 100 bp
10032 10809: contig of 778 bp in length
10810 10909: gap of 100 bp
10910 13046: contig of 2137 bp in length
13047 13146: gap of 100 bp
13147 14912: contig of 1766 bp in length
14913 15012: gap of 100 bp
15013 16787: contig of 1775 bp in length
16788 16887: gap of 100 bp
16888 19027: contig of 2140 bp in length
19028 19127: gap of 100 bp
19128 23490: contig of 4363 bp in length
23491 23590: gap of 100 bp
23591 25951: contig of 2361 bp in length
25952 26051: gap of 100 bp
26052 29248: contig of 3197 bp in length
29249 29348: gap of 100 bp
29349 32751: contig of 3403 bp in length
32752 32851: gap of 100 bp
32852 37464: contig of 4613 bp in length
37465 37564: gap of 100 bp
37565 42592: contig of 5028 bp in length
42593 42692: gap of 100 bp
42693 49347: contig of 6655 bp in length
49348 49447: gap of 100 bp
49448 54633: contig of 5186 bp in length
54634 54733: gap of 100 bp
54734 62288: contig of 7555 bp in length
62289 62388: gap of 100 bp
62389 146838: contig of 84450 bp in length
146939 158635: contig of 11697 bp in length
158636 158735: gap of 100 bp
158736 175143: contig of 16408 bp in length
175144 175243: gap of 100 bp
175244 200086: contig of 24843 bp in length
200087 200186: gap of 100 bp
200187 207214: contig of 7028 bp in length.
Location/Qualifiers
1..207214
```

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-54L14"
/clone_1b="RPCI-23 Female Mouse BAC"
1..9177
/note="assembly_fragment"
vector_end:SP6
vector_side:left"
9278..9931
/note="assembly_fragment"
10032..10809
/note="assembly_fragment"
10910..13046
/note="assembly_fragment"
13147..14912
/note="assembly_fragment"
15013..16787
/note="assembly_fragment"
16888..19027
/note="assembly_fragment"
19128..23490
/note="assembly_fragment"
23591..25951
/note="assembly_fragment"
26052..29248
/note="assembly_fragment"
29349..32751
/note="assembly_fragment"
32852..37464
/note="assembly_fragment"
37565..42592
/note="assembly_fragment"
42693..49347
/note="assembly_fragment"
49448..54633
/note="assembly_fragment"
54734..62288
/note="assembly_fragment"
62389..146838
/note="assembly_fragment"
146939..158635
/note="assembly_fragment"
158736..175143
/note="assembly_fragment"
175244..200086
/note="assembly_fragment"
200187..207214
/note="assembly_fragment"
vector_end:T7
vector_side:right"
BASE COUNT 65311 a 39165 c 39030 g 61707 t 2001 others
ORIGIN
Alignment Scores:
Pred. No.: 0.00292 Length: 207214
Score: 416.00 Matches: 79
Percent Similarity: 90.53% Conservative: 7
Best Local Similarity: 83.16% Mismatches: 9
Query Match: 12.85% Indels: 0
DB: 2 Gaps: 0
US-10-054-935-2 (1-614) x AC107643 (1-207214)
QY 305 GlnProglutenserGluThrSerGlnThrLeuProProlyserPropheserCysGlyArg 324
Db 177968 AGGCCAGGAGATTTGGGGAGATCCAGCTTCTGCTCCAGCCTTTCATGTGCCAG 178027
QY 325 SerGlyLysGlyHisLysArgLysSerProPhesGlySerThrGluArgLysThrProVal 344
Db 178028 AATGGAAAGGACACAAAAGAAACCCATTGGAAATACAGAAAGAAAGAACTCCTGTT 178087
QY 345 LysLysLeuAlaProGluLysSerLysValLysThrLysThrProLysHisSerProIle 364
|||||
```


Db 178088 AAAAGCTGCTCCTGAAATTTTCACAAAGCTCAAAACAAACAAACCTTAAAGCAGCTCTCAGATT 178147
Oy 365 LysgluIupProCysGlySerIenSerGluThrValCysLysArgGluLeuArgSerGln 384
|||||
Db 178148 AAAAGGAAACCTGTGGTTTCATATACAAACTGTTTAAACGGAATTGAGGACCA 178207
Oy 385 GluThrProGluIupProArgSerSerValAspThrProProArg 399
|||||
Db 178208 GAAACCCCAAAAGTCTCGTCTTACAGGAGTATCCACACACAGA 178252
RESULT 9
AC131356 174346 bp DNA linear HTG 21-AUG-2002
LOCUS Rattus norvegicus clone CH230-278A19, *** SEQUENCING IN PROGRESS
DEFINITION *** 55 unordered pieces.
AC131356
AC131356.1 GI:22380622
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS 1 (bases 1 to 174346)
Muzny, D., Marie, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Butrell, R., Calderon, E.,
Cardenas, Y., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., P. Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, A., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gehring, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guerrero, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Hayak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpal, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C.,
Kovals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensushewa, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Maulhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morris, S., Muidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
Newton, N., Nguyen, N., Norris, S., Nwakoelamen, O., Okunolu, G.,
Olaniyinsogun, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldreker, A.,
Popovic, D., Primus, E., Pu, L. L., Puzo, M., Qutro, J., Rachin, E.,
Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbey, A.,
Sisson, I., Sitter, C. D., Smejs, D., Sneed, A., Sodergren, E.,
Song, X. Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
Thigley, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlaczek, R.,
Woods, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhuo, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,
Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
TITLE Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished
2 (bases 1 to 174346)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G00J
Center clone name: CH230-278A19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 109854 bases at least Q40
Consensus quality: 118111 bases at least Q30
Consensus quality: 125098 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1547: contig of 1547 bp in length
1548
1647: gap of unknown length
1648
3064: contig of 1417 bp in length
3065
3164: gap of unknown length
3165
4885: contig of 1721 bp in length
4886
4985: gap of unknown length
4986
6000: contig of 1015 bp in length
6001
6100: gap of unknown length
6101
7171: contig of 1071 bp in length
7172
7271: gap of unknown length
7272
8794: contig of 1523 bp in length
8795
8894: gap of unknown length
8895
10368: contig of 1474 bp in length
10369
10468: gap of unknown length
10469
11906: contig of 1438 bp in length
11907
12006: gap of unknown length
12007
13031: contig of 1025 bp in length
13032
13131: gap of unknown length
13132
14575: contig of 1444 bp in length
14576
14675: gap of unknown length
14676
15699: contig of 1024 bp in length
15700
15799: gap of unknown length
15800
16951: contig of 1152 bp in length
16952
17051: gap of unknown length
17052
18695: contig of 1644 bp in length
18696
18795: gap of unknown length
18796
20470: contig of 1675 bp in length
20471
22418: gap of unknown length
22419
22518: gap of unknown length
22519
24364: contig of 1846 bp in length
24365
24464: gap of unknown length
24465
25586: contig of 1122 bp in length
25587
25686: gap of unknown length
25687
27859: contig of 2173 bp in length
27860
27959: gap of unknown length
27960
29649: contig of 1690 bp in length
29650
29749: gap of unknown length
29750
31331: contig of 1582 bp in length
31332
31431: gap of unknown length

31432	33606	contig of 2175	bp in length
33607	33706	gap of unknown	length
33707	33799	contig of 1473	bp in length
33799	33829	gap of unknown	length
33829	33882	contig of 1914	bp in length
33882	33924	gap of unknown	length
33924	33982	contig of 1089	bp in length
33982	34038	gap of unknown	length
34038	34094	contig of 2277	bp in length
34094	34150	gap of unknown	length
34150	34206	contig of 1904	bp in length
34206	34262	gap of unknown	length
34262	34318	contig of 3136	bp in length
34318	34374	gap of unknown	length
34374	34430	contig of 3040	bp in length
34430	34486	gap of unknown	length
34486	34542	contig of 2523	bp in length
34542	34598	gap of unknown	length
34598	34654	contig of 3045	bp in length
34654	34710	gap of unknown	length
34710	34766	contig of 2207	bp in length
34766	34822	gap of unknown	length
34822	34878	contig of 2985	bp in length
34878	34934	gap of unknown	length
34934	34990	contig of 2541	bp in length
34990	35046	gap of unknown	length
35046	35102	contig of 2034	bp in length
35102	35158	gap of unknown	length
35158	35214	contig of 3116	bp in length
35214	35270	gap of unknown	length
35270	35326	contig of 2106	bp in length
35326	35382	gap of unknown	length
35382	35438	contig of 3267	bp in length
35438	35494	gap of unknown	length
35494	35550	contig of 2909	bp in length
35550	35606	gap of unknown	length
35606	35662	contig of 2243	bp in length
35662	35718	gap of unknown	length
35718	35774	contig of 4198	bp in length
35774	35830	gap of unknown	length
35830	35886	contig of 3666	bp in length
35886	35942	gap of unknown	length
35942	35998	contig of 4944	bp in length
35998	36054	gap of unknown	length
36054	36110	contig of 3448	bp in length
36110	36166	gap of unknown	length
36166	36222	contig of 8106	bp in length
36222	36278	gap of unknown	length
36278	36334	contig of 4933	bp in length
36334	36390	gap of unknown	length
36390	36446	contig of 3841	bp in length
36446	36502	gap of unknown	length
36502	36558	contig of 6548	bp in length
36558	36614	gap of unknown	length
36614	36670	contig of 7435	bp in length
36670	36726	gap of unknown	length
36726	36782	contig of 5906	bp in length
36782	36838	gap of unknown	length
36838	36894	contig of 7286	bp in length
36894	36950	gap of unknown	length
36950	37006	contig of 15613	bp in length

Alignment Scores:	
Pred. NO.:	0.0286
Score:	380.50
Percent Similarity:	27.62%
Best Local Similarity:	22.49%
Query Match:	11.75%
Length:	174346
Matches:	215
Conservative:	49
Mismatches:	183
Indels:	513

DB:	2	caps:	29
US-10-034-935-2 (1-614) x AC131356 (1-174346)			
QY	18	AsnProGluGlnAlaArgLeuAspTyrGluValArgAlaAlaLeuGlyLeuProGluAspGlu	37
Db	56520	AATGATGACCAAGAGG-CAGACACAGAGGAGGAAACAAAGACAGGACAGGAAAGTTCA	56578
QY	38	ProGluAlaAlaGluAlaHisPheLeuProAlaGlnHisArgLeuGlySerGluProGluPro	57
	-----	-----	-----
QY	56579	-----GTAAGAGTCAGGGGCTCCAGAGAAAGTAAAGG----	56608
Db	56609	-----GCCAAGACAGAAAGGCGCAACAGACAGAAAGATGCGAGAGGAAAGGCGAAATG	56652
QY	78	GlyArgGlyLeuLeuLeuProAlaGlyAlaAlaPro-----GlyGln	91
Db	56663	GGAAGAA--CAGATAAAGGAGGAGGCGGAGACAGCCATGTGAACAGAAACAAAGGAGAA	56720
QY	92	GlnGlu-----GluSerTyrGlyGlySerValProLeuProCysProProAla	108
Db	56721	AAGGAAATATACAAAGAAAGATGGGG-----	56747
QY	109	ThyrSerGlnAlaGlyIleGlyGlyGluProAlaAlaAlaGlyAlaGlyCysSerProArg	128
	-----	-----	-----
Db	56748	-----GGGGGGGGCTCTTCGGCCGTG	56768
QY	129	ProTyrTyrGlnAlaValLeuProIleGlnThrGlySerLeuValAlaAlaLysGlu	148
Db	56769	CCTAAG-----	56774
QY	149	ProThrProThrAlaGlyAspLysGlyAlaAlaSerProAlaAlaThrAlaSerAsp	168
Db	56775	TCCACCCCATGGAGTTGGGGGAAGGGGAGTAAATAGGTTCCACGGCTTTTCTCC	56834
QY	169	ProAlaGly-----ProProPro	174
Db	56835	CACCTTAGGTGACTGACATGTAACCTATTATCTGCTATGTGATTAAGAACACACCCCTCT	56894
QY	175	LeuProLeu-Pro-----GlyProProProLeuAl	184
Db	56895	CTCCCTCTTCCCTCTGCTTCTCTACTAGATGGCAGGCTGAAGAGCCCCCACCAATACT	56954
QY	184	AProThrAlaThrAlaGly-----ThrLeuAlaAlaSerGlu	197
Db	56955	GCCCAAGTCCACATCTGCATATGATCTCTGATTCCATACATATGATATCAACAAAGGT	57014
QY	197	YArgTyrLysSerMetArgLysSerProLeuGlyGlyGlyGlySerGlyAlaSer	217
Db	57015	AATGTGGAGA-----GGCCAGGGCGACCTGTAGCC	57047
QY	217	rgln-----AlaAlaCysLeuLysGlnIle--LeuLeuLeuGlnLeuAspLeuIle--	233
Db	57048	TAGACATACAGGGCGTGACAC-AAGCAATAGGCATCTGTGTGCACACGCTGCTGTGTC	57106
QY	234	-----GluGlnGlnGlnGlnGlnLeuGlnAlaLysGluLysGlu	247
Db	57107	AACATGGGTTCCCAACAGTAGCCACCACTAACACAGTGGGTAGGGGCACTATAGGACAT	57166
QY	247	eGluGluLeuLysSerGluArgAspThrLeuLeuAlaArgIleGluArg-----	263
Db	57167	CGAGGACCTGTGTAGCAGCATTTGAGAACTACTGTCATCGATAGCCCGNNNNNNNNNN	57226
QY	263	-----	263
Db	57227	NN	57286
QY	264	-----MetGluArgArgMetGlnLeuValLysLys	274
Db	57287	NNNNNNNNNNNNNNNNNNNNNNNNNNNGTGTTCAAGTTCGACTGCACGTAGTTCTCAAGCA	57346
QY	274	PasGluLysGluArgHisLysLeuPheGlnGlyTyrGluThrGluGluArgGluGlu	294

Db	57347	TTTGGGTAAAGAC	---GGCAGTTTATTATTATTATTTT	-----AC	57382
QY	294	rgIuLeuSerGIuLysIlleuLysleuGluCysGlnProGluLeuSerGIuThrSerGlnTh	314		
Db	57383	TAGTTATGTCAATCATCTATAGTATTTTTTTGAACCTGAGGGAT	-----	57428	
QY	314	rLeuProProLysProPheSerCysGlyArgSerGIuLysGIuLysArgLysSerPr	334		
Db	57428	-----	57428		
QY	334	oPheGIySerThrGIuArgLysThrProValLysLysLeuAlaProGluPheSerLysVa	354		
Db	57429	-----	57429		
QY	354	lYsThrLysThrProLysHisSerProileuLys	---gluGluProCysGlySerLeuSe	373	
Db	57443	GAAACGTACTACATCATGTGCTCATTAAGTAACCTGTCCCTGACCCCTCAAGAACTGGA	57502		
QY	373	rgIuThrValCysLysArgGluLeuArgSerGlnGluThrProGluLysProArgSerSe	393		
Db	57503	TGACAGGTGTCTTTCAAAGAGGCATGCCAAGCTGAGCTAGACAGAGAAAGAGAGAAA	--	57560	
QY	393	rValAspThrProProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGsl	413		
Db	57561	-----	57561		
QY	413	uLysAlaPheSerSerGluIlleGluAspleuProTyLLeuSerThrThrGluMetTyLe	433		
Db	57590	CACTGTGACCCCTGAGACTGTGACTGCCCTCTTCATAC	-----	57626	
QY	433	uGysArgThrHisGlnPro	-----ProProSerProLeuProLeuArgLysLe	449	
Db	57627	-----TGGAGAGTCTCTGGCGCCTAGTCCATAGTA	-----	57659	
QY	449	rSerProLysLysGluGluThrValAlaArgCysLeuMetProSerSerValAlaGlyG1	469		
Db	57660	-----	57660		
QY	469	uThrSerValLeuAlaValProSerThrParGasp	-----	480	
Db	57695	GGCACACATGTGGCCYTT	---TCATGGATGTCCTCCCTCAGTACTACTTGTGTGAAC	57751	
QY	481	-----	481		
Db	57752	ACCAACACAAAATCTAGTGTCTACACCCATCGAGTCTTCC	-----CACCCAGGTAG	57805	
QY	491	-----ProSerAspleuLeuGlnuLeuAspAspSerValPheSerLysArgHisAl	508		
Db	57806	TCAGTTCCTTTTACCCTGAGTAAAGGGTTAAACAAGAACCTTCTAGCCTCGAAGAG	57865		
QY	508	alYsLeuGluLeuAspGluLysArgArg	-----	517	
Db	57866	CAAACTGCTTTGTGA	---GAACACCTTCTAGGTAAAGCTAAATGCTTGTGTGTGTGT	57924	
QY	518	-----LysArgTrpAspIleGlnArgGlyLeuArgGluArgIleLeuGlnAr	533		
Db	57925	TTTGCTGTTTTTCCCATGGATGGATATTCAGAGGATCAGAGACAAAGATTTTACGGG	57984		
QY	533	gluGlnLeuArgMetTyLysLysLysGlyIleGlnGluSerGluProGluValThrSe	553		
Db	57985	ACTGCACCTCGAATGTATTAAGAAAGAAAGAAATTCAGGAATCGACCTGAGAGTTACCC	58044		
QY	553	rPhePheProGluProAspAsp	-----	560	
Db	58045	ATTTTTCCCTGAGCCGATGACGGTAAGTGTATCTATCTAAAAATGTAGAACTTGGGG	58104		
QY	560	-----	560		
Db	58105	GTTGGGCAATTAGCTCAGTGGCAGAGCGCCCTGGGTTCAGTCCCCACGCTCCGAAAAAAG	58164		
QY	560	-----	560		

[illegible]

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Hernandez, O., Hodgson, A., Hugues, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kravovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Soderren, E., Sotake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wlececyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 144577)
 Worley, K.C.

Direct Submission
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 144577)
 Worley, K.C.

Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 18, 2002 this sequence version replaced gi:21240269.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWEG
 Center clone name: CH230-506C22
 ----- Summary Statistics
 Sequencing vector: Plasmid,
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 114009 bases at least Q40
 Consensus quality: 118499 bases at least Q30
 Consensus quality: 121325 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1699: contig of 1699 bp in length

1700 1799: gap of unknown length
 1800 2876: contig of 1077 bp in length
 2877 2976: gap of unknown length
 2977 4489: contig of 1413 bp in length
 4390 4489: gap of unknown length
 4490 5708: contig of 1219 bp in length
 5709 5808: gap of unknown length
 5809 6967: contig of 1159 bp in length
 6968 7067: gap of unknown length
 7068 8623: contig of 1556 bp in length
 8624 8723: gap of unknown length
 8724 9993: contig of 1270 bp in length
 9994 10093: gap of unknown length
 10094 11908: contig of 1815 bp in length
 11909 12009: gap of unknown length
 12009 14387: contig of 2378 bp in length
 14387 14865: gap of unknown length
 14865 16346: contig of 1860 bp in length
 16346 16446: gap of unknown length
 16447 18634: contig of 2187 bp in length
 18634 18733: gap of unknown length
 18734 21914: contig of 3181 bp in length
 21915 22014: gap of unknown length
 22015 24433: contig of 2419 bp in length
 24433 24533: gap of unknown length
 24534 27831: contig of 3298 bp in length
 27831 27931: gap of unknown length
 27931 31839: contig of 3908 bp in length
 31839 31939: gap of unknown length
 31940 33680: contig of 1741 bp in length
 33680 33780: gap of unknown length
 33781 36869: contig of 3089 bp in length
 36870 36969: gap of unknown length
 36970 39041: contig of 2072 bp in length
 39041 39141: gap of unknown length
 39142 42085: contig of 2944 bp in length
 42086 42185: gap of unknown length
 42185 45561: contig of 3376 bp in length
 45561 45661: gap of unknown length
 45662 48095: contig of 2434 bp in length
 48095 48195: gap of unknown length
 48196 53141: contig of 4946 bp in length
 53142 53241: gap of unknown length
 53241 56744: contig of 3503 bp in length
 56745 56844: gap of unknown length
 56845 60010: contig of 3166 bp in length
 60010 60110: gap of unknown length
 60111 64116: contig of 4006 bp in length
 64117 64216: gap of unknown length
 64217 67814: contig of 3598 bp in length
 67815 67914: gap of unknown length
 67915 72543: contig of 4629 bp in length
 72544 72643: gap of unknown length
 72644 75853: contig of 3210 bp in length
 75854 75953: gap of unknown length
 75954 80517: contig of 4564 bp in length
 80518 80617: gap of unknown length
 80618 84461: contig of 3844 bp in length
 84462 84561: gap of unknown length
 84562 89191: contig of 4630 bp in length
 89192 89291: gap of unknown length
 89291 95800: contig of 6509 bp in length
 95801 95900: gap of unknown length
 95901 102912: contig of 7012 bp in length
 102913 103012: gap of unknown length
 103013 107519: contig of 4507 bp in length
 107520 107619: gap of unknown length
 107620 113616: contig of 5997 bp in length
 113617 113716: gap of unknown length
 113717 120950: contig of 7234 bp in length
 120951 121050: gap of unknown length
 121051 127758: contig of 6708 bp in length
 127759 127858: gap of unknown length

* 127859 136315: contig of 8457 bp in length
 * 136316 136415: gap of unknown length
 * 136416 144577: contig of 8162 bp in length.
 Location/Qualifiers
 1.144577
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-506C22"

BASE COUNT 45381 a 25293 c 25234 g 42138 t 6531 others
 ORIGIN

Alignment Scores:

Pred. No.: 6.48 Length: 144577
 Score: 299.00 Matches: 80
 Percent Similarity: 71.09% Conservative: 11
 Best Local Similarity: 62.50% Mismatches: 22
 Query Match: 9.23% Indels: 18
 DB: 2 Gaps: 4

US-10-054-935-2 (1-614) x AC123367 (1-144577)

Oy 354 ValysThrylThrProlys---HisSerProIlelyGluGluProCysGlySerLeu 372
 |||||
 Db 88046 GTCAAAACAAACCTCTACACATTAACTTAAAGAGAACCTGTGGTCCATA 88105
 |||||
 Oy 373 SerGluThrValCysLysArgGluLeuArgSerGluGluThrProGluLysProArgSer 392
 |||||
 Db 88106 TCAGAAACCTTTGTAAACGAGANTGAGGGGCCAGAAAGCCCAAGCCCAAGCTCT 88165
 |||||
 Oy 393 SerValAspThrProProArgLeuSerThrProGluLysGlyProSerThrHisProlys 412
 |||||
 Db 88166 TCAGTGGATACCCACCAAGACTCGAGTCCCA----- 88201
 |||||
 Oy 413 GluLysAlaPheSerSerGluIleGluAspLeuProGlyrLeuSerThrThrGluMetTyr 432
 |||||
 Db 88202 ---AAGCT-CCTCAGTGAAGATGGAAGT-TTGGCCGACCTGTCCACCACTGAATGTAT 88256
 |||||
 Oy 433 LeuCysArgTrpHisGlnProProProSerProLeuArgLeuArgLysSerSerProlys 452
 |||||
 Db 88257 TTATGTCTCTGGGACACCAATTCCTCCGATCACCATTACATTCAGGGGATTCATCTCCA--- 88313
 |||||
 Oy 453 LysGluGluThrValAla-----ArgCysLeuMetProSerSerValAlaGly 468
 |||||
 Db 88314 AAGGAGGAGACGGTGTGACAGTAAGCATAGAGAACATTATATACCC--AGTGTGCCGCT 88371
 |||||
 Oy 469 GluThrSerValLeuAlaValPro 476
 |||||
 Db 88372 CAAGCTAATGTGTCTATGCT 88395
 |||||

RESULT 11

AC107643 207214 bp DNA linear HTG 21-APR-2002
 LOCUS
 DEFINITION Mus musculus clone RP23-54L14, WORKING DRAFT SEQUENCE, 21 ordered
 pieces.

ACCESSION AC107643.2 GI:20258435

VERSION AC107643.2 HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Mus musculus
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 207214)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-54L14
 Unpublished
 2 (bases 1 to 207214)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barran,N., Bastien,V., Boguslavsky,L., Boukhaltier,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collinmore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., Melidim,T., Menus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnuppback,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 207214)

Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barran,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collinmore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Melidim,T., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnuppback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 21, 2002 this sequence version replaced g1:18308342.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRH
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 54.L.14
 Center clone name: 54.L.14

Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version: 0.960731
 Consensus quality: 200414 bases at least Q40
 Consensus quality: 203403 bases at least Q30
 Consensus quality: 204607 bases at least Q20
 Insert size: 21000; agarose-1p
 Insert size: 205214; sum-of-contigs
 Quality coverage: 7.0 in Q20 bases; agarose-1p
 Quality coverage: 7.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have

```

* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1
* 9178 9277: gap of 100 bp
* 9278 9931: contig of 654 bp in length
* 9932 10031: gap of 100 bp
* 10032 10809: contig of 778 bp in length
* 10810 10909: gap of 100 bp
* 10910 13046: contig of 2137 bp in length
* 13047 13146: gap of 100 bp
* 13147 14912: contig of 1766 bp in length
* 14913 15012: gap of 100 bp
* 15013 16787: contig of 1775 bp in length
* 16788 16887: gap of 100 bp
* 16888 19027: contig of 2140 bp in length
* 19028 19127: gap of 100 bp
* 19128 23490: contig of 4363 bp in length
* 23491 23590: gap of 100 bp
* 23591 25951: contig of 2361 bp in length
* 25952 26051: gap of 100 bp
* 26052 29248: contig of 3197 bp in length
* 29249 29348: gap of 100 bp
* 29349 32751: contig of 3403 bp in length
* 32752 32851: gap of 100 bp
* 32852 37464: contig of 4613 bp in length
* 37465 37564: gap of 100 bp
* 37565 42592: contig of 5028 bp in length
* 42593 42692: gap of 100 bp
* 42693 49347: contig of 6555 bp in length
* 49348 49447: gap of 100 bp
* 49448 54633: contig of 5186 bp in length
* 54634 54733: gap of 100 bp
* 54734 62288: contig of 7555 bp in length
* 62289 62388: gap of 100 bp
* 62389 146838: contig of 84450 bp in length
* 146839 146938: gap of 100 bp
* 146939 158635: contig of 11697 bp in length
* 158636 158735: gap of 100 bp
* 158736 175143: contig of 16408 bp in length
* 175144 175243: gap of 100 bp
* 175244 200086: contig of 24843 bp in length
* 200087 200186: gap of 100 bp
* 200187 207214: contig of 7028 bp in length.
FEATURES
source
1. 207214
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1ib="RP23-54LI4"
/clone_1ib="RPCT-23 Female Mouse BAC"
1. 9177
/note="assembly_fragment
clone_end:SP6
vector_side:left"
9278..9931
/note="assembly_fragment"
10032..10809
/note="assembly_fragment"
10910..13046
/note="assembly_fragment"
13147..14912
/note="assembly_fragment"
15013..16787
/note="assembly_fragment"
16888..19027
/note="assembly_fragment"
19128..23490
/note="assembly_fragment"
23591..25951
/note="assembly_fragment"
26052..29248
/note="assembly_fragment"

```

DB	Accession	Gene	Protein	Length	Score	Percent Similarity	Best Local Similarity	Query Match	DB
misc_feature	29349..32751	/note="assembly_fragment"		47.1	274.50	80.56%	77.78%	8.48%	2
misc_feature	32852..37464	/note="assembly_fragment"		274.50	207214	56			
misc_feature	37565..42592	/note="assembly_fragment"		274.50	207214	56			
misc_feature	42693..49347	/note="assembly_fragment"		274.50	207214	56			
misc_feature	49448..54633	/note="assembly_fragment"		274.50	207214	56			
misc_feature	54734..62288	/note="assembly_fragment"		274.50	207214	56			
misc_feature	62389..146838	/note="assembly_fragment"		274.50	207214	56			
misc_feature	146939..158635	/note="assembly_fragment"		274.50	207214	56			
misc_feature	158736..175143	/note="assembly_fragment"		274.50	207214	56			
misc_feature	175244..200086	/note="assembly_fragment"		274.50	207214	56			
misc_feature	200187..207214	/note="assembly_fragment"		274.50	207214	56			
misc_feature	/note="assembly_fragment"			274.50	207214	56			
misc_feature	vector_side:right"			274.50	207214	56			
misc_feature	clone_end:77			274.50	207214	56			
misc_feature	BASE COUNT 65311 a 39165 c 39030 g 61707 t 2001 others			274.50	207214	56			
misc_feature	ORIGIN			274.50	207214	56			
misc_feature	Alignment Scores:			274.50	207214	56			
misc_feature	Pred. No.:			274.50	207214	56			
misc_feature	Score:			274.50	207214	56			
misc_feature	Percent Similarity:			274.50	207214	56			
misc_feature	Best Local Similarity:			274.50	207214	56			
misc_feature	Query Match:			274.50	207214	56			
misc_feature	DB:			274.50	207214	56			
misc_feature	US-10-054-935-2 (1-614) x AC107643 (1-207214)			274.50	207214	56			
misc_feature	390 ProarGserSerValAspThrProProarGleuSerThrProGlnylsglProSerThr			274.50	207214	56			
misc_feature	177976 CCCTGCGCTTCACTGATACCCACCACCAACACTCTGACACCCCAAAAGAGCCACCTCC			274.50	207214	56			
misc_feature	410 HSPolysGlnLysAlaIlePheSerSerGluIleGlnAspLeuProGlyLeuSerThr			274.50	207214	56			
misc_feature	177916 CATCCCAAGAGAAAGCCCTTGAAATGAGATGGAATGCGGACTTTTCATACACA			274.50	207214	56			
misc_feature	430 GluMetTyrLeuCyArGTrpHisGlnProProPheSerProLeuProuleu			274.50	207214	56			
misc_feature	177856 GAAATGTAATTTGTGTTGGCAGACGCTCCGCCATCAACGTTACTGTAGGGTTAGGG			274.50	207214	56			
misc_feature	447 ArgGluSerSerProLysLysGluGlnThrValAla 458			274.50	207214	56			
misc_feature	177976 TTAGAATCCTTCACAAAGAGAGGAGCTGTAGCA 177761			274.50	207214	56			
misc_feature	RESULT 12			274.50	207214	56			
misc_feature	LOCUS CNS080B6 136755 bp DNA linear HTG 12-AUG-2002			274.50	207214	56			
misc_feature	DEFINITION Oryza sativa chromosome 12 clone OSJNB0108014, *** SEQUENCING IN			274.50	207214	56			
misc_feature	ACCESSION AL845344			274.50	207214	56			
misc_feature	VERSION AL845344.1 GI:22208185			274.50	207214	56			
misc_feature	KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.			274.50	207214	56			
misc_feature	SOURCE Oryza sativa.			274.50	207214	56			
misc_feature	ORGANISM Oryza sativa.			274.50	207214	56			
misc_feature	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			274.50	207214	56			
misc_feature	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			274.50	207214	56			

QY	158	ly-----AlaIaIaIaSerProAlaIaIaThrAlaIaSerAspProAlaIaGlyPro	173
Db	672	GAGCGCGGCGGCTGTGAAGCGCGCGCCCGGGGGCCCGGACAGCGCGAGCGCGCGCG	731
QY	173	roPLeuLeuProLeuProGlyProProProLeuAlaProThAlaThrAlaGlyThrLeuA	193
Db	732	CGCGCGCGCGCGCCCGAGAGCGGCTCCGGGGGGGCGGCATGGCTCGGGGCGGCGG	790
QY	193	IaIaIaSerGluGlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGly---	211
Db	791	CGCGCGCGCGCTCGAGG-----ACGAACCTGTGGCCGGGCGGCGGCGGAGAGA	841
QY	212	--GlySer-----	213
Db	842	GAGGCTCCGAGCGGGTACCGGACCCGTCGCCCGGAGCGCCCGCGCGAGCGCGC	901
QY	214	-----GlyAlaSerSerGlnAlaIaIaCysLeuLysGlnIleLeuLeuGlnI	230
Db	902	CGAATATTGACATGGGCGGAGCGGAGCGGCGG-----	933
QY	230	euaSpleuIleGluGlnGlnGlnGlnGlnLeuGlnAlaIaLysGluLysGluIleGluI	250
Db	934	-----CGGCGCGGAGCCGAGCGCGGAGCGCGGAGCGCGGAG--GAGAGAG	972
QY	250	eulYsSerGluArgAspThrLeuLeuAlaIaGlyIleGluArgMetLuarArgMetGln	270
Db	973	AGTCGGGCGCGCGGACCATGTGTCCAGCCGGTGAGCCAGCTCAAGAGACCGAGAACG	1033
QY	270	euValYsLysAspAsnGluLysGlnLuarGlnIaLysLeuPheGlnIaYyrGluThrGlu	290
Db	1033	CTTTCATGCTGTGTGGCGGCGAGCGCGCGGAGATGGCCAG-----GAGAACCCA	1088
QY	290	IuaArgGluIuThrGluLeuSerLysLysIleLysLeuGluLysGlnProGluLeuSerG	310
Db	1087	AGATGCACAACATCGAGATGACAGAACGGCGTGCGGCGCGGAGTGAAG--CHCTGAGGG	1144
QY	310	IuThrSerGlnIuThrLeuProLysProPhe--SerCysGlyArgSerGlyLysGlnI	329
Db	1144	AGGCGGAGAG--CGGCTTCATCTGACAGAGGCCAAGGCGGTGCGGGCGCA	1191
QY	330	Lys-----ArgLysSerProPheGlySerThrGluArgLysThrProValLysLysLeu	347
Db	1192	TGCACATGAAAGAGACACCCGACATCAAGTACACCGCGCGCGCGCA-----	1233
QY	348	AlaProGluPheSerLysValLysThrLysThrProLysHisSerProIleLysGluIu	367
Db	1237	AGCCCAAGACGCTCTCAAGAAAGCAAGTTCGCTCCGCTGACGAGCGGTGGGG	1299
QY	368	ProCysGlySerLeuSerGluThrVal-----CysLysArgGluLeu	381
Db	1297	CCGTGGCGACACAGAGGCCCCCGACGGGTGAAGGCGGCGCGCTGACAGCGCGGGCTG	1358
QY	382	ArgSerGlnIuThrProGluLys-----ProArgSerSerValAspThrPro	397
Db	1357	CGGCGGGGCTGGGCGCCACATCGCTGCTGGGCACACCAGAGAGGCGGCGCGCGCG	1411
QY	398	ProArgLeuSerThrProGlnLysGlyProSerThrHisProLysGluIuLysAlaPheSer	417
Db	1417	CCG-----CGCGCGCGCGCGCGCTTCT	1444
QY	418	SerGluIleGluAspLeuProLysLeuSerThrThrGluMetLysCysArgTrpHis	437
Db	1441	TCCCGCAGATCGCGCGCGCG-----	1465
QY	438	GlnProProProSerProLeuProLeuArgLysSer-----	449
Db	1462	CGGCGCGCGCGCGCGCGCGCGCGGGGGGCGCCCTACTCGCTGTGACCTGGGCT	1522
QY	450	-----SerProLysLysGluIuThrValAlaArgCysLeuMetProSer	464
Db	1522	CCAAGATGCGCGAGATCTCTCTGTGTGTGGGGGCTGGCGGCTGCCCTACGCTCTCT	1588
QY	465	SerValAlaGlyGluThrSerValLeuAlaValProSerTrpArgAspHisSerValGlu	484

Db	Accession	Source	Length	Score
1582	CGCTGGGCT-----ACCCGCGCGCGCGCGCGGAGACCTTCACGCGG	1626	10.1	3385
QY	485 ProLeuArgAspProAsnPro 491			
Db	1627 CCGCGCGCGCGCGCGCGCGCG 1647			
RESULT 14				
AB020652	3385 bp	linear	PRI 16-JUN-1999	
LOCUS	Homo sapiens mRNA for KIAA0845 protein, partial cds.			
DEFINITION	AB020652			
VERSION	AB020652.1			
SOURCE	GI:4240178			
ORGANISM	Homo sapiens adult male brain cDNA to mRNA, clone_1lb:pBluescriptII SK plus clone:hk05234.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (sites)			
TITLE	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirotsawa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.			
JOURNAL	Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro			
REFERENCE	DNA Res. 5 (6), 355-364 (1998)			
AUTHORS	2 (bases 1 to 3385)			
TITLE	Ohara,O., Suyama,M., Kikuno,R., Nagase,T. and Ishikawa,K.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (02-DEC-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba			
AUTHORS	292-0812, Japan (E-mail:cdna1nfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)			
FEATURES	Location/Qualifiers			
SOURCE	1..3385			
gene	/organism="Homo sapiens"			
CDS	/db_xref="taxon:9606"			
	/clone="hk05234"			
	/sex="male"			
	/tissue_type="brain"			
	/clone_1lb="pBluescriptII SK plus"			
	/dev_stage="adult"			
	1..3385			
	/gene="KIAA0845"			
	<1..2802			
	/gene="KIAA0845"			
	/codon_start=1			
	/product="KIAA0845 protein"			
	/protein_id="BAA74868.1"			
	/db_xref="GI:4240179"			
	/translation="SRSEKQLOALNDFPAGYIDKYNROLEAHNRSLGEAAALRQOQA GRAMGLYREVREMGAVLRICGAARGLRQEHLEDTIAHYRORLDEARREERA EAAARALARRQGEAREARVDLQKKAQLQEBGCTLRHHDEVEVELGQIDGSSAQA QMAETDALKCDVTSALREIRQLQEGHAAVQSTQSEEWFRVDRDISEAKVNTDMA RSQOETTEYRROLQARTTELEALKSKDSLEORSLEDRHQADIASYQDAIOQLDA ELNRTKEMAAOLREYQDILNVKALDIEIAAYRKLLEGECRIGFQIPSLPEGIP KITSVSHIKVISEEKIKVVESEKIEVIIEBOETQVTEVEEERKEAKEEGEKK EEGGEAEAGGGEEMTSPAAEASPEKAKSPVKEAKSPAKKSPKEAKSPAE VKSPEKASPAKEAKSPPEAKSPPEKAKSPVAKVSPPEKASPAKEAKSPAKKSP EKASPVKEAKSPAEAKSPVKEAKSPAEKSPPEKASPAKEAKSPPEKASPAK SPEKEAKSPPEKASPAKASPEKASPAKASPEKASPAKASPEKASPAKASPA VKEAKSPPEKASPAKASPEKASPAKASPEKASPAKASPEKASPAKASPAK AKAEARSPADKPEKASPAKASPEKASPAKASPEKASPAKASPEKASPAK EKQEDVVKPEKASPAKASPEKASPAKASPEKASPAKASPEKASPAKASPA KESVKEAKKEAEADKKVPTPEKASPAKASPEKASPAKASPEKASPAKASPA PAEKKEAPKKDTPKEKAKPEKPEKTEAKAKEDDKTLSKSPKPAEKASSSSTD QKDSKPEKAKTEKAAAGK"			
BASE COUNT	1051 a	869 c	1037 g	428 t
ORIGIN				

Score: 246.00 Matches: 181
Percent Similarity: 31.53% Conservative: 81
Best Local Similarity: 21.78% Mismatches: 291
Query Match: 7.60% Gaps: 278
DB: 9 Gaps: 32

US-10-054-935-2 (1-614) x AB020652 (1-3385)

```

QY 11 ALAAlaAlaProAlaGlyGlyAsnProGluGlnArgLeuAspTryGluAlaAlaAla 30
DB 114 GTCGCGGACAGCAGCAGCGCGG-----CCGCTCCGCTATGGCGAGCGTACGA 161
QY 31 LeuGlyGlyProGluAsp---GluProGlyAlaAlaGluAlaHisPheLeuProArgHis 49
DB 162 GCGCGAGGTCTCCGAGATGCGCGCGCGCGCTCGCTCGCGTGG-----CGCGCG 209
QY 50 ArgLysLeuLysGluPro-----GlyProProLeuAlaSerSerGlnGlyLysPro 67
DB 210 GCGCGGACACTGACCTGGAGCAGAGACCTGCTCGAGAGACATCGCGACGTCGCCA 269
QY 68 AlaPro-----SerProAlaGlyCysGlyGlyLysGlyArgGlyLeuLeuPro 84
DB 270 GCGCGTACAGCAGAGAGCGCGCGCAGAGAGAGCGCGCGCGCGCGCGCTGGC 329
QY 85 Ala-----GlyAlaAlaProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 97
DB 330 GCGCTTCGCGCAGAGCGCGCGCGCGCGCGCGCTGACCTGACGAGAGAGCGCGCT 389
QY 98 ---GlySerValProLeuProCysProProGluAlaThrLysGlnAlaGlyIleGlyGly 116
DB 390 GCGAGGAGAGTGGCGCTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 422
QY 117 GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTryGlnAlaValLeuPro 136
DB 423 ---CGAGGAGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467
QY 137 IleGlnThrGlySerLeuValAlaAlaAlaLysGluProThrProThrAlaGlyAspLys 156
DB 468 ---CGCGCGCGCAGCGCGAGATGCAAGCGCGAGCGCGCGCGCGCGCGCGCGCG 524
QY 157 GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProLeuPro 176
DB 525 GTCGCGCGCTGCGGAGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 584
QY 177 LeuProGlyProProProLeuAlaProAlaThrAlaGlyThrLeuAlaAlaSerGlu 196
DB 585 GTCCGAGGA-----GTGGTTCGAGTGAAGCGTGGACCGACTGTCGAGGAGCGCA 635
QY 197 GlyArgTrpLysSerMetArgLysSerProLeuGlyGlyGly----- 211
DB 636 GGT-----GAACACAGACGCTATGCGCTCAGCGCAGGAGAGATAGTACCG 686
QY 211 -----GAACACAGACGCTATGCGCTCAGCGCAGGAGAGATAGTACCG 211
DB 687 GCGTCACTGACAGCGCAGACCCAGACCTGAGCGCAGTAAGAACCCAGAGACTCACT 746
QY 212 -----GlySerGlyAlaSerSerGlnAlaAlaLysLeu-LysGlnIle 225
DB 747 GAGAGGACAGCGCTGAGCTGAGGAGCGCTCATACAGCGCAGCATTCCTCTACACGA 806
QY 225 eLeuLeuLeuGlnLeuAspLeu---IleGluGlnGlnGlnGlnGlnGlnGlnGlnGln 244
DB 807 AGCCATTCAACACGCTGAGCTGAGAGAACACCAAGTGGAGATGCGCGCGCGCGCT 866
QY 244 uLysGluIleGluGlnLeu----- 250
DB 867 GCGAGATACACAGGACCTGCTCAATGTCAGATGCGCTTGATATAGATAGAGTACCGCTTA 926
QY 250 ----- 250
DB 927 CAGAAAACCTCTGGAAGTGAAGAGTGTGATTCGCTTGGCCAAATCTCTTCGCT 986
QY 250 ----- 250

```

```

DB 987 TCCAGAGAGACTCCCAAAATTCCTCTGTGTCCACTCACATTAAGGTGAAGCGAGAGA 1046
QY 251 -----LysSerGluArgAspThrLeuLeuAlaArgIleGluArgme 264
DB 1047 GAAGATCAAGTGTGTGAGAAAGTCTTGAAAGAACTGTGATTT-----GTGAGAGACA 1100
QY 264 tGluArgArgMetGlnLeuVal-----LysLysAspAsnGlu 277
DB 1101 GACAGAGAGACCCAAAGTCACTGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1160
QY 277 sGluArgHisLysLeuPheGlnGlyTrpGluThrGlnGluArgGlu----- 292
DB 1161 GAGAGAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 1220
QY 293 -----GluThrGluLeuSerGluLysIle 300
DB 1221 AACCAATCTCCCGCAGACAGAGAGCTGATCCCGCAGAGAGAGAGAGAGAGAGAG 1280
QY 300 eLysLeuGluCysGln----- 305
DB 1281 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
QY 305 ----- 305
DB 1341 CCCAGCCGAAAGTCAAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1400
QY 306 -----ProGluLeuSerGluThrSerGlnThrLeuProProLysPr 319
DB 1401 ACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1460
QY 319 oPheSerCysGlyArgSerGlyLysGlnLysLysArgLysSerProPheGlnSerThrGln 339
DB 1461 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
QY 339 uArgLysThrProValLysLysLeuAlaProGluPheSerLysValLysThrLysThrPr 359
DB 1512 GGCCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1571
QY 359 oLysHisSerProIleLysGlnGlu-----ProCysGln 370
DB 1572 GCGCAAGTCCCGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631
QY 370 ySerLeuSerGluThrValCysLysArgGluLeuArgSerGlnGlu----- 385
DB 1632 GCGCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
QY 386 -----ThrProGluLysProArgSer 393
DB 1686 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
QY 393 rVal-----AspThrProProArgLeuSerThrProGlnLysGlyProSerTh 409
DB 1746 AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
QY 409 rHisPro-----LysGluLysAlaPheSerSerGlnIleGluAs 422
DB 1806 GTCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1865
QY 422 pLeuProTrpLysSerThrThrGlnMetLysLysLysLysLysLysLysLysLys 442
DB 1866 GTCCCGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1904
QY 442 rProLeuProLeuArgGlu-----SerSerProLysLysGln 454
DB 1905 CAAGTCCCGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964
QY 454 uGluThrValAlaArgCysLeuMetProSerSerValAlaGlyLysThrSerValLeuAl 474
DB 1965 AGAAGCCAAAGTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2024
QY 474 aValProSerThrArgAspHisSerValGluProLeuArgAspProAsn-----Pr 491

```

[illegible]

BASE COUNT	1119 a	996 c	1172 g	487 t
ORIGIN				
Alignment Scores:				
Pred. No.:	11	Length:	3774	
Score:	246.00	Matches:	181	
Percent Similarity:	31.53%	Conservative:	81	
Best Local Similarity:	21.78%	Mismatches:	291	
Query Match:	7.60%	Indels:	278	
DB:	9	Gaps:	32	
US-10-054-935-2 (1-614) x AF203032 (1-3774)				
QY	11	ALAAlaAlaProAlaGlyGlyAlaSerProGluGlnArgLeuAspTyrGluAlaAlaAla	30	
DB	393	GCTGGCGAGACAGAGAGCGGGG-----CCGCTCCGCTATGGCGCGCTGTACGA	440	
QY	31	LeuGlyGlyProGluAsp-----GluProGlyAlaAlaGluAlaHisPheLeuProArgHis	49	
DB	441	GCGCGAGGCTCCGCGAGATGCCGGCGCGCTCTCGCTCGG-----CGGCGC	488	
QY	50	ArgTyrLeuLysGluPro-----GlyProProLeuAlaSerSerGlnGlySerPro	67	
DB	489	GCGCGCTAGCTACGCGCTTGGAGCAGAGACACTCTCTCGAGACATCCGCGACGTGGCGCA	548	
QY	68	AlaPro-----SerProAlaGlyCysGlyGlyLysGlyArgGlyLeuLeuLeuPro	84	
DB	549	GCGCTTACACAGCAGAGCCCGGACGAGAGAGAGCGCGGCGCGCGCTGGC	608	
QY	85	Ala-----GlyAlaAlaProGlyGlnGlnGluGlnSerTyrProGly---	97	
DB	609	GCGCTTGGCAGAGAGCGCGAGCGCGCGCGCTGGACCTGCAGAAAGAGCGCAGCGCT	668	
QY	98	---GlySerValProLeuProCysArgProProProProAlaThrLysGlnAlaGlyIleGlyGly	116	
DB	669	GCAGGAGGAGTGGCGGTACTCTGCGCGCGCAC-----	701	
QY	117	GluProAlaAlaAlaGlyAlaGlyCysSerProArgProLysTyrGlnAlaValLeuPro	136	
DB	702	---CCAGGAAAGAGTGGCGAGCTGCTGGCGAGATCCAGGCGTCCGG-----	746	
QY	137	IleGlnThrGlySerLeuValAlaAlaAlaLysGluProThrProTyrAlaGlyLys	156	
DB	747	---CCGCCGAGCGCGCAGATGCAGAGCCGAGACGCGCCCTGGAAGTCCGACGTGAC	803	
QY	157	GlyGlyAlaAlaSerProAlaAlaThrAlaSerAspProAlaGlyProProProLeuPro	176	
DB	804	GTCGGCGCTGGCGGAGTTCGCGCGACCTTGAAAGCCAGCGGTGCAGACAGCTGCA	863	
QY	177	LeuProGlyProProProLeuAlaProThrAlaThrAlaGlyThrLeuAlaAlaSerGlu	196	
DB	864	GTCGAGGA-----GTGGTTCGAGTGAAGCTGACGCGAGCTGTGGAGAGCA	914	
QY	197	GlyArgTyrLysSerMetArgLysSerProLeuGlyGlyGlyGly	211	
DB	915	GGT-----GAACACAGAGCGTATCGCTCAGCGAGAGAGATTAAGTACG	965	
QY	211	-----	211	
DB	966	GCGTAGTCGACGCGCAGACACAGAGACTGGAGGCACTGAAAGACCAAGCACTACT	1025	
QY	212	-----GlySerGlyAlaSerSerGlnAlaAlaCysLeu-LysGlnIle	225	
DB	1026	GGAGAGCGAGCGCTGTGAGTGGAGAGCCGTCATCAGGCCAGACATATGCTCTTACCGAGA	1085	
QY	225	eleuLeuLeuGlnLeuAspLeu---IleGluGlnGlnGlnGlnLeuGlnAlaLysGlu	244	

